Agrégation d'états dans les processus markoviens

Gerardo Rubino, Bruno Sericola

UNITÉ DE RECHERCHE
INRIA-RENNES

Rapports de Recherche

N° 858

AGREGATION D'ETATS DANS LES PROCESSUS MARKOVIENS

Gerardo RUBINO
Bruno SERICOLA

JUIN 1988

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 36 20 00
Télex : UNIRISA 950 473 F
Télécopie : 99 38 38 32

Publication Interne n° 401
AVRIL 1988
48 PAGES

Agrégation d'états dans les processus markoviens

Gerardo Rubino et Bruno Sericola

1 avril 1988

Résumé Nous analysons les conditions sous lesquelles le processus agrégé par rapport à une partition fixée de l'espace d'état (fini) d'un processus markovien homogène est encore un processus markovien homogène. On traite d'abord le cas à temps discret en donnant une caratérisation de l'agrégation d'états, puis on traite le cas à temps continu en montrant qu'il est toujours possible de se ramener au cas discret par uniformisation.

PROCESSUS MARKOVIENS, AGREGATION, AGREGATION FAIBLE, UNIFORMISATION

State aggregation in Markov processes

Abstract We analyse under which conditions the aggregated process with respect to a given partition of the (finite) state space of an homogeneous Markov process is still an homogeneous Markov process. We first consider the discrete time case giving a characterization of the state aggregation and then we consider the continuous time one showing that it is always possible to come down to the discrete time case using uniformization.

MARKOV PROCESSES, AGGREGATION, WEAK LUMPABILITY, UNIFORMIZATION

1 Introduction

Les processus markoviens sont à la base des modèles analytiques dans le domaine de l'évaluation quantitative des systèmes. En particulier, dans le cadre des systèmes informatiques, ils sont largement utilisés pour résoudre des problèmes de performance ou de fiabilité.

Dans de nombreuses situations, l'utilisateur n'a pas besoin d'informations sur chacun des états de son modèle mais sur des agrégats ou classes d'états. Cela peut se produire, par exemple, lorsque les paramètres intéressants sont directement associés à des sous-ensembles de l'espace d'état (celui-ci étant parfois trop important pour que des informations sur les états eux-mêmes soient significatives). On peut rencontrer aussi ce genre de situations lorsque l'on augmente l'espace d'état d'un modèle, en ajoutant des états fictifs, pour obtenir un modèle markovien. En outre, de manière générale, plus une modélisation est fine, plus le nombre d'états générés est important (réseaux de Petri stochastiques, réseaux de files d'attente, ...). Enfin, une situation analogue se produit lorsque le système étudié n'est observable que sur une partition de l'espace d'état. Ces considérations peuvent nous amener à construire un nouveau processus, dit *agrégé*, dont les états sont des classes d'états du processus originel.

Dans ce contexte, une question importante qui se pose est de savoir si le processus agrégé est aussi un processus de Markov homogène. Dans l'affirmative on conserve toute la puissance des outils d'analyse des processus de Markov mais malheureusement ce n'est pas généralement le cas. Le but de ce rapport est d'analyser les conditions que doit satisfaire le processus originel pour que l'agrégé soit markovien homogène.

Il est bien connu que si une certaine propriété (voir Théorème 3.15 de ce rapport) est vérifiée par la matrice des probabilités de transition d'une chaîne de Markov homogène à temps discret alors le processus agrégé est une chaîne de Markov homogène. On a le même résultat dans le cas continu en ce qui concerne la matrice des taux de transition (voir Théorème 4.20 de ce rapport). Il est moins bien connu que cette condition n'est pas nécéssaire: il existe des processus de Markov homogènes tels que le processus agrégé par rapport à une certaine partition de l'espace d'état est aussi markovien homogène sans que la propriété en question soit vérifiée. Dans [3], Kemeny et Snell ont étudié cette situation qu'ils ont appelé *agrégation faible*. C'est essentiellement l'existence de ces processus qu'ils ont montré; les auteurs donnent aussi une condition suffisante pour obtenir l'agrégation faible. Ce travail a été repris et développé dans [1] puis dans [6]. Dans [1], une méthode est proposée pour aborder le problème de la caractérisation de ces processus; cette méthode est reprise et étendue dans [6] pour obtenir des nouvelles propriétés ainsi qu'une première caractérisation de l'agrégation faible. Ces résultats concernent seulement le cas des processus de Markov à temps discret.

Dans ce rapport, on donne une caractérisation des processus de Markov faiblement agrégeables sous la forme d'un algorithme ayant un nombre borné de pas ou itérations. Ceci constitue un raffinement des résultats présentés dans [6]. De plus, le cas continu est aussi étudié et on montre que l'on peut se ramener au cas discret par uniformisation du processus originel. Dans le but de présenter des résultats qui puissent être exploitables dans des applications, c'est à dire qui permettent raisonnablement d'aboutir dans les calculs concernés, et pour limiter la place occupée par les développements, nous traitons seulement le cas des processus à espace d'état fini.

L'organisation de ce rapport est la suivante. Après quelques généralités et notations qui font l'objet de la Section 2, la Section 3 traite le cas discret (on parlera de chaînes de Markov) en passant d'abord en revue les résultats précédemment connus et en les complétant. Dans la Section 4, on étudie le cas continu (on parlera de processus markoviens) en montrant que le problème peut toujours être ramené au cas discret par uniformisation. Des conclusions font l'objet de la dernière section.

2 Généralités et notations

On considère un processus markovien homogène X, que l'on suppose irréductible, à temps discret ou continu. La distribution stationnaire de ce processus X sera notée π. L'espace d'état est fini et on le note $E = \{1, 2, ..., N\}$. On note $\mathcal{B} = \{B(1), B(2), ..., B(M)\}$ une partition de l'espace d'état E et $n(i)$ le cardinal de la classe $B(i)$. Pour plus de clarté, on suppose les états de E ordonnés de la façon suivante:

$$\begin{array}{lcl} B(1) & = & \{1, \ldots, n(1)\} \\ B(2) & = & \{n(1)+1, \ldots, n(1)+n(2)\} \\ & \vdots & \\ B(m) & = & \{n(1)+ \ldots + n(m-1)+1, \ldots, n(1)+ \ldots + n(m)\} \\ & \vdots & \\ B(M) & = & \{n(1)+ \ldots + n(M-1)+1, \ldots, N\} \end{array}$$

Au processus X, on associe le processus stochastique agrégé Y sur l'espace d'état $F = \{1, 2, ..., M\}$, defini par:

$$Y_t = m \iff X_t \in B(m) \text{ pour tout } t.$$

On déduit aisément de cette définition et de l'irréductibilité de X que le processus Y obtenu est aussi irréductible en ce sens que $\forall m \in F$, $\forall l \in F$ tel que $P(Y_0 = l) > 0$, $\exists t \geq 0$ tel que $P(Y_t = m / Y_0 = l) > 0$.

On notera $\mathcal{A}$ l'ensemble des vecteurs de probabilités de taille N. Par convention, tous les vecteurs utilisés sont des vecteurs lignes. Les vecteurs colonnes seront représentés par l'opérateur $(.)^T$. Un vecteur dont toutes les composantes sont égales à 1 sera noté simplement 1, sa dimension étant définie par le contexte.

Pour $\alpha \in \mathcal{A}$ et $B \in \mathcal{B}$, on notera α^B le vecteur de $\mathcal{A}$ défini par:

$$\alpha^B(i) = \begin{cases} \dfrac{\alpha(i)}{\sum_{j \in B} \alpha(j)} & \text{si } i \in B \\ 0 & \text{sinon} \end{cases}$$

D'autre part, pour tout $l \in F$, on note $\alpha_{B(l)}$ la *restriction* de α à la classe $B(l)$: ce vecteur ligne a pour dimension $n(l)$ et sa $k^{ième}$ composante est la $(n(1) + \ldots + n(l-1) + k)^{ième}$ composante de α. La fonction $\alpha \mapsto \alpha_{B(l)}$ sera notée T_l, c'est à dire, $T_l.\alpha = \alpha_{B(l)}$. Réciproquement, pour tout vecteur de probabilité β de dimension $n(l)$, $T_l^{-1}.\beta = \gamma \in \mathcal{A}$ avec $\gamma(i) = 0$ si $i \notin B(l)$ et $\gamma(i) = \beta(i)$ si $i \in B(l)$. On remarquera en particulier que si $\alpha_{B(l)} \neq 0$, on a:

$$\alpha^{B(l)} = \frac{T_l^{-1}.\alpha_{B(l)}}{\alpha_{B(l)} 1^T} \tag{1}$$

On peut alors énoncer les deux lemmes suivants:

Lemme 2.1 *Tout élément $\alpha \in \mathcal{A}$ s'écrit de manière unique comme combinaison convexe des $\alpha^{B(l)}$ lorsqu'ils sont définis. C'est à dire:*

$$\alpha = \sum_{\substack{l \in F, \\ \alpha_{B(l)} \neq 0}} \alpha_{B(l)} 1^T \alpha^{B(l)}$$

Démonstration. Soit $\alpha \in \mathcal{A}$, on a:

$$\begin{aligned} \alpha &= \sum_{l \in F} T_l^{-1}.\alpha_{B(l)} \\ &= \sum_{\substack{l \in F, \\ \alpha_{B(l)} \neq 0}} \alpha_{B(l)} 1^T \alpha^{B(l)} \quad \text{par la relation (1)} \end{aligned}$$

Pour l'unicité, si

$$\alpha = \sum_{\substack{l \in F, \\ \alpha_{B(l)} \neq 0}} \lambda_l \alpha^{B(l)} \quad \text{avec } \sum_{l \in F} \lambda_l = 1,\ 0 \leq \lambda_l \leq 1$$

alors pour tout $l \in F$ tel que $\alpha_{B(l)} \neq 0$ on a: $\alpha_{B(l)} 1^T = \lambda_l \alpha^{B(l)} 1^T = \lambda_l$ □

Voici quelques exemples. Prenons $N = 5$ et $\mathcal{B} = \{B(1), B(2)\}$ avec $B(1) = \{1,2,3\}$ et $B(2) = \{4,5\}$. Si $\alpha = (1/10, 1/10, 1/5, 2/5, 1/5)$ on a:

$$\begin{array}{ll} \alpha^{B(1)} = (1/4, 1/4, 1/2, 0, 0) & \alpha^{B(2)} = (0, 0, 0, 2/3, 1/3) \\ \alpha_{B(1)} = (1/10, 1/10, 1/5) & \alpha_{B(2)} = (2/5, 1/5) \end{array}$$

$$\begin{array}{l} T_1.\alpha^{B(1)} = (1/4, 1/4, 1/2) \\ T_2^{-1}.(1/4, 3/4) = (0, 0, 0, 1/4, 3/4) \end{array}$$

pour $\alpha = (1/2, 1/2, 0, 0, 0)$

$$\alpha^{B(1)} = (1/2, 1/2, 0, 0, 0) \quad \alpha^{B(2)} \text{n'est pas défini} \quad (\alpha_{B(2)} = (0,0))$$

Lemme 2.2 *Soit $l \in F$, $x_1, \ldots, x_n$ n vecteurs de $\mathcal{A}$, $\lambda_1, \ldots, \lambda_n$ n réels ≥ 0 tels que $\sum_{i=1}^n \lambda_i = 1$ et K une constante positive quelconque, alors:*

$$\left(K \sum_{i=1}^n \lambda_i x_i\right)^{B(l)} = \left(\sum_{i=1}^n \lambda_i x_i\right)^{B(l)} = \sum_{\substack{i/1 \leq i \leq n, \\ (x_i)_{B(l)} \neq 0}} \frac{\lambda_i (x_i)_{B(l)} 1^T}{\sum_{j=1}^n \lambda_j (x_j)_{B(l)} 1^T} (x_i)^{B(l)}$$

$$\textit{lorsque } \left(\sum_{i=1}^n \lambda_i x_i\right)^{B(l)} \textit{ est défini, c'est à dire,}$$

$$\textit{lorsque } \left(\sum_{i=1}^n \lambda_i x_i\right)_{B(l)} \neq 0$$

Démonstration. La première égalité est évidente par l'effet de la normalisation. Quant à la deuxième, en développant, on obtient:

$$\begin{aligned} \left(\sum_{i=1}^n \lambda_i x_i\right)^{B(l)} &= \frac{T_l^{-1}.\left(\sum_{i=1}^n \lambda_i x_i\right)_{B(l)}}{\left(\sum_{i=1}^n \lambda_i x_i\right)_{B(l)} 1^T} \quad \text{par (1)} \\ &= \frac{T_l^{-1}.\sum_{i=1}^n \lambda_i (x_i)_{B(l)}}{\sum_{i=1}^n \lambda_i (x_i)_{B(l)} 1^T} \\ &= \frac{\sum_{i=1}^n \lambda_i T_l^{-1}.(x_i)_{B(l)}}{\sum_{i=1}^n \lambda_i (x_i)_{B(l)} 1^T} \\ &= \sum_{\substack{i/1 \leq i \leq n, \\ (x_i)_{B(l)} \neq 0}} \frac{\lambda_i (x_i)_{B(l)} 1^T}{\sum_{j=1}^n \lambda_j (x_j)_{B(l)} 1^T} (x_i)^{B(l)} \quad \text{par (1)} \end{aligned}$$

□

Une notation qui sera constamment utilisée dans la suite est la suivante: dans des égalités de la forme $P_u(f(X)) = P_v(g(X))$, X est une famille de processus de Markov de probabilités de transition fixées (ou de taux de transition fixés dans le cas continu). La partie gauche est la probabilité de l'événement $f(X)$ pour le processus de Markov de la famille dont la distribution initiale est u; la partie droite fait référence au processus X muni de la distribution initiale v $(u, v \in \mathcal{A})$.

3 Le cas discret

Soit $X = (X_n)_{n \geq 0}$ une chaîne de Markov homogène sur l'espace d'état E. On suppose que la partition $\mathcal{B}$ est fixée. X est donnée par sa matrice des probabilités de transition P et par sa distribution de probabilité initiale α. On notera (α, P) cette chaîne de Markov lorsque ce sera nécessaire. On suppose X irréductible et les éléments de P seront notés $P(i, j)$. On notera $agg(\alpha, P, \mathcal{B})$ la chaîne agrégée construite à partir de (α, P) par rapport à la partition $\mathcal{B}$.

On considérera souvent la famille de toutes les chaînes de Markov sur le même espace d'état E ayant la même matrice des probabilités de transition P. Cette famille sera notée $(\,.\,, P)$. Rappelons qu'en général, la distribution de probabilité initiale figurera en indice de la mesure de probabilité. Par exemple, $P_\beta(X_n \in B)$ représente la probabilité qu'après n transitions, la chaîne de Markov (β, P) se trouve dans un des états du sous-ensemble B de E (la matrice des probabilités de transition P étant bien évidemment fixée).

On notera aussi $P(i, B)$ la probabilité de passer en une transition de l'état i à l'un des états du sous-ensemble B de E, c'est à dire, $P(i, B) = \sum_{j \in B} P(i, j)$. Si l'on considère la décomposition de la matrice P en blocs relativement à la partition $\mathcal{B}$, alors on notera $P_{B(i)B(j)}$ le bloc $n(i) \times n(j)$ correspondant aux transitions de $B(i)$ vers $B(j)$.

Tout d'abord, nous présentons quelques lemmes que nous référencerons fréquemment par la suite.

3.1 Préliminaires

Lemme 3.1 *Soit B, $D \in \mathcal{B}$,*

$$\begin{array}{lrcl} (i) & P_\alpha(X_n \in B) & = & P_{\alpha P^n}(X_0 \in B) \\ (ii) & P_\alpha(X_n \in B / X_0 \in D) & = & P_{\alpha^D}(X_n \in B) \\ (iii) & P_\alpha(X_{n+1} \in B / X_n \in D) & = & P_{\alpha P^n}(X_1 \in B / X_0 \in D) \end{array}$$

Démonstration.

Remarquons tout d'abord que $P_\beta(X_0 \in B) = \sum_{i\in B} \beta(i)$. Maintenant:

$$(i)\quad P_\alpha(X_n \in B) \;=\; \sum_{i\in B} \pi_n(i) \quad \text{où } \pi_n = \alpha P^n \text{ représente la distribution de } X_n$$
$$= P_{\pi_n}(X_0 \in B)$$

$$\begin{aligned}
(ii)\quad P_\alpha(X_n \in B / X_0 \in D) &= \frac{\sum_{d\in D} \alpha(d) P^n(d,B)}{\sum_{d\in D} \alpha(d)} \\
&= \sum_{d\in D} \left(\frac{\alpha(d)}{\sum_{d'\in D} \alpha(d')} \right) P^n(d,B) \\
&= P_{\alpha^D}(X_n \in B)
\end{aligned}$$

$$\begin{aligned}
(iii)\quad P_\alpha(X_{n+1} \in B / X_n \in D) &= \frac{\sum_{d\in D} P_\alpha(X_n = d) P(d,B)}{P_\alpha(X_n \in D)} \\
&= \sum_{d\in D} \left(\frac{P_{\alpha P^n}(X_0 = d)}{P_{\alpha P^n}(X_0 \in D)} \right) P(d,B) \quad \text{par (i)} \\
&= \sum_{d\in D} P_{(\alpha P^n)^D}(X_0 = d) P(d,B) \\
&= P_{(\alpha P^n)^D}(X_1 \in B) \\
&= P_{\alpha P^n}(X_1 \in B / X_0 \in D) \quad \text{par (ii)}
\end{aligned}$$

□

Dans toute la suite, nous utiliserons fréquemment la définition suivante:

Définition 3.2 *Une suite $(C_0, C_1, \ldots, C_j)$ de sous-ensembles de E sera dite possible pour la distribution initiale α si $P_\alpha(X_0 \in C_0, X_1 \in C_1, \ldots, X_j \in C_j) > 0$. En particulier, si $B \in \mathcal{B}$, (B) est une suite possible pour α si $\alpha_B \neq 0$.*

Définition 3.3 *Soit $\alpha \in \mathcal{A}$ et $(C_0, C_1, \ldots, C_j)$ une suite d'éléments de $\mathcal{B}$ possible pour α, on définit le vecteur $f(\alpha, C_0, C_1, \ldots, C_j) \in \mathcal{A}$ récursivement par:*

$$\begin{aligned}
f(\alpha, C) &= \alpha^C \\
f(\alpha, C_0, C_1, \ldots, C_k) &= (f(\alpha, C_0, C_1, \ldots, C_{k-1}) P)^{C_k}
\end{aligned}$$

Par exemple, $f(\alpha, B, C, D) = ((\alpha^B P)^C P)^D$.

Lemme 3.4 *La restriction de $\beta = f(\alpha, C_0, C_1, \ldots, C_j)$ au sous-ensemble C_j est donnée par:*

$$\beta_{C_j} = K^{-1}\alpha_{C_0}P_{C_0C_1}P_{C_1C_2}\ldots P_{C_{j-1}C_j}$$
$$\text{où } K = \alpha_{C_0}P_{C_0C_1}P_{C_1C_2}\ \ldots P_{C_{j-1}C_j}1^T$$

Démonstration. La propriété est claire pour $j = 0$. Supposons qu'elle soit vraie pour toute suite ayant n sous-ensembles et notons $\gamma = f(\alpha, C_0, C_1, \ldots, C_{n-1})$. On a alors:

$$\gamma_{C_{n-1}} = L^{-1}\alpha_{C_0}P_{C_0C_1}P_{C_1C_2}\ldots P_{C_{n-2}C_{n-1}}$$
$$\text{où } L = \alpha_{C_0}P_{C_0C_1}P_{C_1C_2}\ldots P_{C_{n-2}C_{n-1}}1^T.$$

Si $\beta = f(\alpha, C_0, C_1, \ldots, C_n)$ alors $\beta = (\gamma P)^{C_n}$ par définition. Donc,

$$\beta_{C_n} = H^{-1}(\gamma P)_{C_n} \quad \text{où } H = (\gamma P)_{C_n}1^T$$

Mais, $(\gamma P)_{C_n} = \gamma_{C_{n-1}}P_{C_{n-1}C_n}$ puisque $\gamma(i) = 0$ si $i \notin C_{n-1}$. D'où,

$$\begin{aligned} \beta_{C_n} &= H^{-1}\gamma_{C_{n-1}}P_{C_{n-1}C_n} \\ &= H^{-1}L^{-1}\alpha_{C_0}P_{C_0C_1}P_{C_1C_2}\ldots P_{C_{n-2}C_{n-1}}P_{C_{n-1}C_n} \end{aligned}$$

Ce qui donne le résultat attendu puisque K^{-1} est simplement une constante de normalisation.

□

Lemme 3.5 *Soit $(C_0, \ldots, C_j)$ une suite d'éléments de $\mathcal{B}$.*

Soit $\mathcal{A}(C_0, \ldots, C_j) \stackrel{\text{déf}}{=} \{\gamma \in \mathcal{A} /$ la suite $(C_0, \ldots, C_j)$ est possible pour $\gamma\}$.

$\mathcal{A}(C_0, \ldots, C_j)$ est convexe et s'il est non vide, on a la propriété suivante:

$$\forall \alpha, \beta \in \mathcal{A}(C_0, \ldots, C_j),\ \forall \lambda, \mu \geq 0 \ \textit{ tels que } \lambda + \mu = 1,$$

$$f(\lambda\alpha + \mu\beta, C_0, \ldots, C_j) = \lambda\frac{K_\alpha}{K}f(\alpha, C_0, \ldots, C_j) + \mu\frac{K_\beta}{K}f(\beta, C_0, \ldots, C_j)$$
$$\textit{où } K_\alpha = \alpha_{C_0}P_{C_0C_1}\ldots P_{C_{j-1}C_j}1^T,$$
$$K_\beta = \beta_{C_0}P_{C_0C_1}\ldots P_{C_{j-1}C_j}1^T$$
$$\textit{et } K = \lambda K_\alpha + \mu K_\beta.$$

Démonstration. Soit B un sous-ensemble de E; pour tout α, $\beta \in \mathcal{A}$ et pour tout λ, $\mu \in \mathbb{R}$, on a:

$$(\lambda\alpha + \mu\beta)_B = \lambda\alpha_B + \mu\beta_B$$

Soit $\alpha, \beta \in \mathcal{A}(C_0, \ldots, C_j)$ et $\lambda, \mu \geq 0$ avec $\lambda + \mu = 1$; soit K_α, K_β et K comme dans l'énoncé. Par le Lemme 3.4, $\alpha \in \mathcal{A}(C_0, \ldots, C_j) \Longleftrightarrow K_\alpha > 0$. On a donc $K_\alpha > 0$ et $K_\beta > 0$. Ensuite,

$$(\lambda\alpha + \mu\beta)_{C_0} P_{C_0C_1} P_{C_1C_2} \ldots P_{C_{j-1}C_j} 1^T = \lambda K_\alpha + \mu K_\beta > 0$$

donc, $\mathcal{A}(C_0, \ldots, C_j)$ est convexe.

Enfin,

$$\begin{aligned} f(\lambda\alpha + \mu\beta, C_0, \ldots, C_j) &= T_j^{-1}.K^{-1}(\lambda\alpha + \mu\beta)_{C_0} P_{C_0C_1} P_{C_1C_2} \ldots P_{C_{j-1}C_j} \\ &= K^{-1}\lambda T_j^{-1}.\alpha_{C_0} P_{C_0C_1} P_{C_1C_2} \ldots P_{C_{j-1}C_j} \\ &+ K^{-1}\mu T_j^{-1}.\beta_{C_0} P_{C_0C_1} P_{C_1C_2} \ldots P_{C_{j-1}C_j} \\ &= \lambda \frac{K_\alpha}{K} f(\alpha, C_0, \ldots, C_j) + \mu \frac{K_\beta}{K} f(\beta, C_0, \ldots, C_j) \end{aligned}$$

□

Dans le lemme fondamental suivant, nous généralisons le Lemme 3.1 (iii).

Lemme 3.6 *Pour tout $k \geq 0$ on a:*

$$P_\alpha(X_{n+1} \in B / X_n \in C_n, \ldots, X_{n-k} \in C_{n-k}) = P_\beta(X_1 \in B) \quad \textit{pour tout } n \geq k$$

$$\textit{où } \beta = f(\alpha P^{n-k}, C_{n-k}, \ldots, C_n)$$

Démonstration. On démontre ce résultat par récurrence sur le nombre de sous-ensembles C_j. S'il n'y a qu'un sous-ensemble, la démonstration est identique à celle du Lemme 3.1 (iii). Supposons que le résultat soit vrai pour l sous-ensembles C_j, $l \leq k$, alors:

$$\begin{aligned} & P_\alpha(X_{n+1} \in B / X_n \in C_n, \ldots, X_{n-k} \in C_{n-k}) \\ = \; & \frac{P_\alpha(X_{n+1} \in B,\ X_n \in C_n, \ldots, X_{n-k} \in C_{n-k})}{P_\alpha(X_n \in C_n, \ldots, X_{n-k} \in C_{n-k})} \\ = \; & \frac{\sum_{c \in C_n} P_\alpha(X_{n+1} \in B / X_n = c) P_\alpha(X_n = c / X_{n-1} \in C_{n-1}, \ldots, X_{n-k} \in C_{n-k})}{\sum_{c \in C_n} P_\alpha(X_n = c / X_{n-1} \in C_{n-1}, \ldots, X_{n-k} \in C_{n-k})} \\ = \; & \sum_{c \in C_n} P(c, B) \left(\frac{P_\gamma(X_1 = c)}{P_\gamma(X_1 \in C_n)} \right) \quad \text{où } \gamma = f(\alpha P^{n-k}, C_{n-k}, \ldots, C_{n-1}) \\ & \quad \text{(par l'hypothèse de récurrence)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{c\in C_n} P(c,B)\left(\frac{P_{\gamma P}(X_0=c)}{P_{\gamma P}(X_0\in C_n)}\right) \quad \text{(par le Lemme 3.1 (i))}\\
&= \sum_{c\in C_n} (\gamma P)^{C_n}(c)P(c,B)\\
&= P_\beta(X_1\in B) \ \text{ puisque } (\gamma P)^{C_n}=\beta
\end{aligned}
$$

□

Le résultat suivant donne une interprétation probabiliste de la fonction f.

Lemme 3.7 *Le vecteur $\beta = f(\alpha, C_0, C_1, \ldots, C_j)$ est la distribution de X_j sachant $\{X_0 \in C_0, \ldots, X_j \in C_j\}$. C'est à dire:*

$$\beta(i) = P_\alpha(X_j = i/X_0 \in C_0, \ldots, X_j \in C_j)$$

Démonstration. Si $i \notin C_j$ le résultat est trivial. Soit $i \in C_j$,

$$
\begin{aligned}
& P_\alpha(X_j = i/X_0 \in C_0, \ldots, X_j \in C_j)\\
&= \frac{P_\alpha(X_j = i,\ X_0 \in C_0, \ldots, X_j \in C_j)}{P_\alpha(X_0 \in C_0, \ldots, X_j \in C_j)}\\
&= \frac{P_\alpha(X_j = i, X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})}{P_\alpha(X_0 \in C_0, \ldots, X_j \in C_j)}\\
&= \frac{P_\alpha(X_j = i/X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})P_\alpha(X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})}{P_\alpha(X_j \in C_j/X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})P_\alpha(X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})}\\
&= \frac{P_\alpha(X_j = i/X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})}{P_\alpha(X_j \in C_j/X_0 \in C_0, \ldots, X_{j-1} \in C_{j-1})}\\
&= \frac{P_\delta(X_1 = i)}{P_\delta(X_1 \in C_j)} \quad \text{où } \delta = f(\alpha, C_0, \ldots, C_{j-1}) \text{, grâce au Lemme 3.6}\\
&= (\delta P)^{C_j}(i) = \beta(i)
\end{aligned}
$$

d'où le résultat énoncé.

□

Enfin le dernier résultat de cette section est un lemme technique.

Lemme 3.8 *La fonction $v \mapsto P_v(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k})$ pour j, k, $C_{j-1}, \ldots, C_{j-k}$, B fixés, est continue sur $\mathcal{A}$.*

Démonstration. Posons $g(v) = P_v(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k})$. Alors:

$$g(v) = \sum_{i \in E,\, v(i)>0} v(i) P_v(X_j \in B / X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k},\ X_0 = i)$$

Mais $P_v(X_j \in B / X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k},\ X_0 = i)$ a la même valeur pour tout $v \in \mathcal{A}$ tel que $v(i) > 0$. On peut donc écrire:

$$\begin{aligned}
&|g(v') - g(v)| \\
&= \Big| \sum_{\substack{i \in E, \\ v(i) > 0,\ v'(i) > 0}} (v'(i) - v(i)) P(X_j \in B / X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k},\ X_0 = i) \Big| \\
&\leq \sum_{\substack{i \in E, \\ v(i) > 0,\ v'(i) > 0}} |v'(i) - v(i)| P(X_j \in B / X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k},\ X_0 = i) \\
&\leq \sum_{i \in E} |v'(i) - v(i)| \\
&= \|v' - v\|
\end{aligned}$$

ce qui implique la continuité de g sur $\mathcal{A}$. □

3.2 Agrégation forte et agrégation faible

Nous analysons maintenant la chaîne $Y = agg(\alpha, P, \mathcal{B})$. Pour tout $B \in \mathcal{B}$, on notera $\mathcal{A}(\alpha, B)$ le sous-ensemble de $\mathcal{A}$ des distributions de probabilité de la forme $f(\alpha, C_1, \ldots, B)$, c'est à dire:

$$\mathcal{A}(\alpha, B) \overset{\text{déf}}{=} \{\ \beta \in \mathcal{A} \ /\ \exists j \geq 0 \text{ et une suite } (C_1, \ldots, C_j,\ B) \text{ possible pour } \alpha, \text{ réduite à } (B) \text{ si } j = 0, \text{ tels que } \beta = f(\alpha,\ C_1, \ldots, C_j,\ B)\ \}$$

Il est facile de vérifier que pour tout $\alpha \in \mathcal{A}$ et pour tout $B \in \mathcal{B}$, l'ensemble $\mathcal{A}(\alpha, B)$ n'est pas vide; c'est une conséquence de l'irréductibilité de Y. Notons aussi que si B est réduit à un seul état i, alors le cardinal de $\mathcal{A}(\alpha, B)$ est égal à 1 puisque $\mathcal{A}(\alpha, \{i\}) = \{e_i\}$ où e_i est la distribution initiale ayant un 1 à la $i^{\text{ième}}$ position. Avec cette notation, une première caractérisation du fait que Y est une chaîne de Markov homogène est donnée par le théorème suivant:

Théorème 3.9 ([3]) *La chaîne $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène si et seulement si $\forall l, m \in F$, la probabilité $P_\beta(X_1 \in B(m))$ a la même valeur pour tout $\beta \in \mathcal{A}(\alpha, B(l))$. Cette valeur commune est la probabilité de transition de l'état l vers l'état m pour la chaîne Y.*

Démonstration. La démonstration est essentiellement basée sur le Lemme 3.6. Ce lemme nous permet d'écrire:
$\forall l, m \in F$, $\forall n \geq 0$ et $\forall (C_0, C_1, \ldots, C_{n-1}, B(l))$ possible pour α,

$$P_\alpha(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1},\ \ldots, X_0 \in C_0) = P_\beta(X_1 \in B(m))$$

$$\text{où } \beta = f(\alpha,\ C_0,\ C_1,\ \ldots, C_{n-1},\ B(l))$$

Si la condition du théorème est satisfaite, la quantité ci-dessus ne dépend que de l et m, et ceci est équivalent au fait que Y est une chaîne de Markov homogène.

Réciproquement, supposons que Y est une chaîne de Markov homogène. Soit β un vecteur quelconque de $\mathcal{A}(\alpha, B(l))$. Le vecteur β peut donc s'écrire:
$\beta = f(\alpha,\ C_0,\ C_1,\ \ldots, C_{n-1},\ B(l))$. En réutilisant le Lemme 3.6, on écrit:

$$\begin{aligned} P_\beta(X_1 \in B(m)) &= P_\alpha(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1},\ \ldots, X_0 \in C_0) \\ &= P_\alpha(X_{n+1} \in B(m)/X_n \in B(l)) \end{aligned}$$

qui ne dépend pas de n par homogénéité de Y, donc, $P_\beta(X_1 \in B(m))$ a la même valeur pour tout $\beta \in \mathcal{A}(\alpha, B(l))$.
□

Un cas particulier est donné par le corollaire suivant:

Corollaire 3.10 *Si pour tout $B, D \in \mathcal{B}$ $(\pi^B P)^D = \pi^D$ alors $Y = agg(\pi, P, \mathcal{B})$ est une chaîne de Markov homogène.*

Démonstration. Si cette condition est satisfaite, l'ensemble $\mathcal{A}(\pi, D)$ est réduit au seul élément π^D, ce qui entraîne immédiatemment le résultat via le théorème précédent.
□

Plus généralement, de la même manière, on a: si pour tout $B, D \in \mathcal{B}$ tels que α^B, α^D et $(\alpha^B P)^D$ sont définis, $(\alpha^B P)^D = \alpha^D$ alors $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène.

Nous allons étudier de façon générale l'ensemble des distributions initiales α conduisant à une chaîne de Markov homogène pour $Y = agg(\alpha, P, \mathcal{B})$. Nous noterons cet ensemble $\mathcal{A}_M$, c'est à dire:

$$\mathcal{A}_M \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A} / Y = agg(\alpha, P, \mathcal{B}) \text{ est une chaîne de Markov homogène}\}$$

Le cas $\mathcal{A}_M \neq \emptyset$ et $\mathcal{A}_M \neq \mathcal{A}$ existe comme nous le montre l'exemple suivant [3].

Soit la famille de chaînes de Markov $(. , P)$ où

$$P = \left(\begin{array}{c|cc} 1/4 & 1/4 & 1/2 \\ \hline 0 & 1/6 & 5/6 \\ 7/8 & 1/8 & 0 \end{array} \right)$$

et soit $\mathcal{B} = \{B(1), B(2)\}$, $B(1) = \{1\}$ et $B(2) = \{2, 3\}$. Considérons les vecteurs de la forme $\alpha_a = (1 - 3a, a, 2a)$ lorsque a décrit l'intervalle $[0, 1/3]$. On peut alors remarquer que:

- $\alpha_a P = \alpha_b$ avec $b = (3 - 4a)/12$
- $(\alpha_a)^{B(1)} = \alpha_0$
- $(\alpha_a)^{B(2)} = \alpha_{1/3}$

On obtient donc:
$\mathcal{A}(\alpha_a, B(1)) = \{\alpha_0\} = \{(1, 0, 0)\}$ et $\mathcal{A}(\alpha_a, B(2)) = \{\alpha_{1/3}\} = \{(0, 1/3, 2/3)\}$
La condition du Théorème 3.9 est donc trivialement satisfaite pour toute distribution initiale de la forme α_a. En d'autres termes, $\{\alpha_a,\ a \in [0, 1/3]\} \subseteq \mathcal{A}_{\mathcal{M}}$.

Prenons maintenant comme distribution initiale le vecteur $\alpha = (0, 0, 1)$. Posons $\beta = f(\alpha, B(2), B(2))$ et $\gamma = f(\beta, B(2), B(2))$. Les vecteurs β et γ appartiennent à $\mathcal{A}(\alpha, B(2))$. Un calcul simple nous donne:

$$\beta = (0, 1, 0) \text{ et } \gamma = (0, 1/6, 5/6)$$

Or,

$$P_\beta(X_1 \in B(1)) = 0 \text{ et } P_\gamma(X_1 \in B(1)) = 35/48$$

Donc, la distribution initiale α ne conduit pas à une chaîne de Markov homogène pour Y, i.e. $(0, 0, 1) \notin \mathcal{A}_{\mathcal{M}}$.

On peut aussi vérifier que la condition du Corollaire 3.10 est satisfaite (la distribution stationnaire est $\pi = (7/16, 3/16, 6/16)$).

Nous allons maintenant énoncer quelques propriétés de l'ensemble $\mathcal{A}_{\mathcal{M}}$.

Théorème 3.11 *Si $\alpha \in \mathcal{A}_{\mathcal{M}}$ alors:*

- $\alpha P^j \in \mathcal{A}_{\mathcal{M}}$ *pour tout* $j \in \mathbb{N}$
- $\alpha^{B(k)} \in \mathcal{A}_{\mathcal{M}}$ *pour tout $k \in F$ tel que la suite $(B(k))$ soit possible pour α*

Démonstration. Soit $j \in \mathbb{N}$. Par un argument similaire à celui utilisé pour la démonstration du Lemme 3.6, il est simple de vérifier que:

$$P_\alpha(X_{n+j+1} \in B/X_{n+j} \in D,\ X_{n+j-1} \in C_{n-1}, \ldots, X_j \in C_0)$$
$$= P_{\alpha P^j}(X_{n+1} \in B/X_n \in D,\ X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0)$$

Par hypothèse, la partie gauche de cette égalité est indépendante de n et de la suite $(C_0,\ C_1,\ \ldots,\ C_{n-1})$ considérée (les classes B et D étant fixées). Donc, $\alpha P^j \in \mathcal{A}_\mathcal{M}$ pour tout $j \in \mathbb{N}$.

Pour prouver le deuxième point, prenons $\alpha \in \mathcal{A}_\mathcal{M}$ et $k \in F$ avec $\alpha_{B(k)} \neq 0$. On a alors:

$$P_{\alpha^{B(k)}}(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1}, \ldots, X_0 \in B(k)) = P_\delta(X_1 \in B(m))$$

$$\text{où } \delta = f(\alpha^{B(k)},\ B(k), \ldots, C_{n-1},\ B(l)) = f(\alpha,\ B(k), \ldots, C_{n-1},\ B(l))$$

Donc,

$$P_\delta(X_1 \in B(m)) = P_\alpha(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1}, \ldots, X_0 \in B(k))$$

qui a la même valeur pour tout $\delta \in \mathcal{A}\left(\alpha^{B(k)}, B(l)\right)$, ce qui prouve que $\alpha^{B(k)} \in \mathcal{A}_\mathcal{M}$.
□

Cette propriété nous permet de prouver le lemme technique suivant.

Lemme 3.12 *Si $\alpha \in \mathcal{A}_\mathcal{M}$ alors $\frac{1}{n}\sum_{k=1}^{n} \alpha P^k \in \mathcal{A}_\mathcal{M}$*

Démonstration. Posons

$$\gamma_n = \frac{1}{n}\sum_{k=1}^{n} \alpha P^k$$

On a par le Lemme 3.6:

$$P_{\gamma_n}(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0) = P_{\gamma_n'}(X_1 \in B(m))$$

$$\text{où } \gamma_n' = f(\gamma_n,\ C_0,\ \ldots,\ C_{n-1},\ B(l))$$

Si $\alpha_k' = f(\alpha P^k,\ C_0,\ \ldots,\ C_{n-1},\ B(l))$, on a par le Lemme 3.5:

$$\gamma_n' = \frac{1}{n}\sum_{k=1}^{n} \frac{K_k}{K}\alpha_k'$$

où $K_k = (\alpha P^k)_{C_0} P_{C_0C_1} \dots P_{C_{n-1}B(l)} \mathbb{1}^T$ et $K = \frac{1}{n}\sum_{k=1}^{n} K_k$

On obtient alors

$$P_{\gamma'_n}(X_1 \in B(m)) = \frac{1}{n}\sum_{k=1}^{n} \frac{K_k}{K} P_{\alpha'_k}(X_1 \in B(m))$$

Or:

$$\begin{aligned} P_{\alpha'_k}(X_1 \in B(m)) &= P_{\alpha P^k}(X_{n+1} \in B(m)/X_n \in B(l), \dots, X_0 \in C_0) \\ &= P_\alpha(X_{n+k+1} \in B(m)/X_{n+k} \in B(l), \dots, X_k \in C_0) \\ &= P_\alpha(X_{j+1} \in B(m)/X_j \in B(l)) \\ & \quad \forall j \text{ tel que } P_\alpha(X_j \in B(l)) > 0, \text{ puisque } \alpha \in \mathcal{A}_{\mathcal{M}} \end{aligned}$$

Finalement, pour tout j tel que $P_\alpha(X_j \in B(l)) > 0$ on a:

$$\begin{aligned} P_{\gamma'_n}(X_1 \in B(m)) &= \frac{1}{n}\sum_{k=1}^{n} \frac{K_k}{K} P_\alpha(X_{j+1} \in B(m)/X_j \in B(l)) \\ &= P_\alpha(X_{j+1} \in B(m)/X_j \in B(l)) \end{aligned}$$

qui est donc constant sur $\mathcal{A}(\gamma_n, B(l))$, ce qui signifie: $\gamma_n \in \mathcal{A}_{\mathcal{M}}\ \forall n \geq 1$. □

Le théorème suivant prouve l'unicité de la matrice des probabilités de transition de la chaîne de Markov homogène agrégée.

Théorème 3.13 *Si $\mathcal{A}_{\mathcal{M}} \neq \emptyset$ et si $\hat{P}$ désigne la matrice des probabilités de transition de la chaîne de Markov homogène agrégée $Y = agg(\alpha, P, \mathcal{B})$ alors la matrice $\hat{P}$ est la même pour tout α donnant une chaîne de Markov homogène agrégée et de plus $\pi \in \mathcal{A}_{\mathcal{M}}$*

Démonstration. Soit $\alpha \in \mathcal{A}$ tel que $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène (i.e. $\alpha \in \mathcal{A}_{\mathcal{M}}$).

On a par le Lemme 3.12 que

$$\alpha \in \mathcal{A}_{\mathcal{M}} \Longrightarrow \frac{1}{n}\sum_{j=1}^{n} \alpha P^j \in \mathcal{A}_{\mathcal{M}} \text{ pour tout } n \geq 1$$

Désignons par $\hat{P}(l, m)$ la probabilité de transition de l vers m pour Y. On a alors:

$$\forall k \text{ tel que } P_\alpha(X_k \in B(l)) > 0,\quad \begin{aligned}[t] \hat{P}(l, m) &= P_\alpha(X_{k+1} \in B(m)/X_k \in B(l)) \\ &= P_{\alpha P^k}(X_1 \in B(m)/X_0 \in B(l)) \end{aligned}$$

Soit n suffisament grand pour que $\sum_{j=1}^{n}(\alpha P^j)_{B(l)}1^T > 0$ (un tel n existe car X est irréductible). Pour simplifier l'écriture, posons pour $k = 1, \ldots, n$:

$$\gamma_k = \frac{(\alpha P^k)_{B(l)}1^T}{\sum_{j=1}^{n}(\alpha P^j)_{B(l)}1^T}$$

$\hat{P}(l,m)$ peut alors s'écrire de la façon suivante:

$$\begin{aligned}
\hat{P}(l,m) &= \sum_{k=1}^{n} \gamma_k \hat{P}(l,m) \qquad \text{car } \sum_{j=1}^{n} \gamma_j = 1 \\
&= \sum_{\substack{k/1 \le k \le n, \\ P_\alpha(X_k \in B(l)) > 0}} \gamma_k P_{\alpha P^k}(X_1 \in B(m)/X_0 \in B(l))
\end{aligned}$$

que l'on peut encore écrire par le Lemme 3.1 (ii)

$$\begin{aligned}
\hat{P}(l,m) &= \sum_{\substack{k/1 \le k \le n, \\ P_\alpha(X_k \in B(l)) > 0)}} \gamma_k P_{(\alpha P^k)^{B(l)}}(X_1 \in B(m)) \\
&= P_\Gamma(X_1 \in B(m))
\end{aligned}$$

$$\text{où } \Gamma = \sum_{\substack{k/1 \le k \le n, \\ P_\alpha(X_k \in B(l)) > 0)}} \gamma_k (\alpha P^k)^{B(l)}$$

En utilisant le Lemme 2.2, Γ peut s'écrire:

$$\Gamma = \left(\sum_{k=1}^{n} \alpha P^k\right)^{B(l)} = \left(\frac{1}{n}\sum_{k=1}^{n} \alpha P^k\right)^{B(l)}$$

On obtient finalement:

$$\begin{aligned}
\hat{P}(l,m) &= P_{(\frac{1}{n}\sum_{k=1}^{n} \alpha P^k)^{B(l)}}(X_1 \in B(m)) \\
&= P_{\frac{1}{n}\sum_{k=1}^{n} \alpha P^k}(X_1 \in B(m)/X_0 \in B(l))
\end{aligned}$$

En faisant tendre n vers $+\infty$ et grâce au Lemme 3.8, on obtient:

$$\hat{P}(l,m) = P_\pi(X_1 \in B(m)/X_0 \in B(l))$$

qui ne dépend pas de α. □

Définition 3.14 *On dira que la famille de chaînes de Markov $(., P)$ est faiblement agrégeable par rapport à la partition $\mathcal{B}$ si et seulement si l'ensemble $\mathcal{A}_{\mathcal{M}} \neq \emptyset$. Pour tout $\alpha \in \mathcal{A}_{\mathcal{M}}$, la chaîne agrégée $Y = agg(\alpha, P, \mathcal{B})$ sera alors une chaîne de Markov homogène et on notera $\hat{P}$ sa matrice des probabilités de transition qui est la même pour tout $\alpha \in \mathcal{A}_{\mathcal{M}}$ par la Théorème 3.13.*

Un cas particulier important est le suivant. Si pour toute paire de classes D, B de la partition $\mathcal{B}$, $P(d, B)$ a la même valeur pour tout $d \in D$, alors la condition du Théorème 3.9 est vérifiée. En effet, soit β un vecteur de $\mathcal{A}(\alpha, \mathcal{D})$: $\beta = f(\alpha, C_0, C_1, \ldots, C_{n-1}, D)$. On a:

$$\begin{aligned}
P_\beta(X_1 \in B) &= P_\alpha(X_{n+1} \in B / X_n \in D,\ X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0) \\
&= \frac{\sum_{d \in D} P(d, B) P_\alpha(X_n = d / X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0)}{\sum_{d \in D} P_\alpha(X_n = d / X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0)} \\
&\quad \text{(Lemme 3.6)} \\
&= P(d, B) \text{ puisque } P(d, B) \text{ a la même valeur pour tout } d \in D
\end{aligned}$$

Dans ce cas, pour tout $\alpha \in \mathcal{A}$, $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène. On dira alors que la famille $(., P)$ est *agrégeable* ou *fortement agrégeable* par rapport à la partition $\mathcal{B}$. La réciproque est aussi vraie comme l'établit le théorème suivant:

Théorème 3.15 ([3]) $(., P)$ *est fortement agrégeable par rapport à la partition $\mathcal{B}$ si et seulement si pour toute paire de sous-ensembles $D, B \in \mathcal{B}$, $P(d, B)$ a la même valeur pour tout $d \in D$. Cette valeur commune est la probabilité de transition de D vers B pour la chaîne Y.*

Démonstration. Nous venons de voir que la condition est suffisante.

Réciproquement, si pour toute distribution $\alpha \in \mathcal{A}$, $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène alors c'est en particulier vrai pour la distribution initiale α_d dont la composante d de D vaut 1. En appliquant maintenant le Théorème 3.9, on voit que $P_\beta(X_1 \in B)$ a la même valeur pour tout $\beta \in \mathcal{A}(\alpha_d, D)$ et vaut $P_{\alpha_d}(X_1 \in B / X_0 \in D)$ De plus, le Théorème 3.13 nous dit que cette valeur est indépendante de la distribution initiale, donc de d; or cette valeur est exactement égale à $P(d, B)$. $\square$

Une autre propriété importante de l'ensemble $\mathcal{A}_{\mathcal{M}}$ est donnée par le théorème suivant:

Théorème 3.16 *L'ensemble $\mathcal{A}_{\mathcal{M}}$ est un convexe fermé.*

Démonstration. Par le Lemme 3.8, il est immédiat de voir que l'ensemble $\mathcal{A}_{\mathcal{M}}$ est un fermé. Pour prouver la convexité de cet ensemble, supposons que α et $\beta \in \mathcal{A}_{\mathcal{M}}$ et posons $\gamma = \lambda\alpha + \mu\beta$ où λ, $\mu \geq 0$ avec $\lambda + \mu = 1$.

Par le Lemme 3.6, on a:

$$P_\gamma(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1}, \ldots, X_0 \in C_0) = P_{\gamma'}(X_1 \in B(m))$$

$$\text{où } \gamma' = f(\gamma,\ C_0, \ldots, C_{n-1},\ B(l))$$

Si l'on note $\alpha' = f(\alpha,\ C_0, \ldots, C_{n-1},\ B(l))$ et $\beta' = f(\beta,\ C_0, \ldots, C_{n-1},\ B(l))$, le Lemme 3.5 nous permet d'écrire:

$$\gamma' = \lambda\frac{K_\alpha}{K}\alpha' + \mu\frac{K_\beta}{K}\beta'$$

$$\text{Où } K_\alpha = \alpha_{C_0}P_{C_0C_1}\ldots P_{C_{j-1}C_j}1^T,\ K_\beta = \beta_{C_0}P_{C_0C_1}\ldots P_{C_{j-1}C_j}1^T$$
$$\text{et } K = \lambda K_\alpha + \mu K_\beta$$

On obtient alors:

$$\begin{aligned}
P_{\gamma'}(X_1 \in B(m)) &= P_{(\lambda K_\alpha\alpha' + \mu K_\beta\beta')/K}(X_1 \in B(m)) \\
&= \lambda\frac{K_\alpha}{K}P_{\alpha'}(X_1 \in B(m)) + \mu\frac{K_\beta}{K}P_{\beta'}(X_1 \in B(m)) \\
&= \lambda\frac{K_\alpha}{K}\hat{P}(l,m) + \mu\frac{K_\beta}{K}\hat{P}(l,m) \text{ par le Théorème 3.13} \\
&= \hat{P}(l,m)
\end{aligned}$$

ce qui prouve que $\gamma \in \mathcal{A}_{\mathcal{M}}$ et donc que l'ensemble $\mathcal{A}_{\mathcal{M}}$ est convexe. □

Introduisons une nouvelle notation pour désigner l'enveloppe convexe formée à partir des points $\pi^{B(i)}$, $i \in F$.

$$\mathcal{A}_\pi = \{\sum_{i\in F} \lambda_i \pi^{B(i)},\ \lambda_i \geq 0 \text{ et } \sum_{i\in F} \lambda_i = 1\}$$

Corollaire 3.17 *Si l'ensemble $\mathcal{A}_{\mathcal{M}}$ est non vide alors $\mathcal{A}_\pi \subseteq \mathcal{A}_{\mathcal{M}}$.*

Démonstration. La démonstration est une conséquence immédiate des Théorèmes 3.11, 3.13, 3.16. □

3.3 Caractérisation de l'agrégation faible

De manière à caractériser l'ensemble $\mathcal{A}_{\mathcal{M}}$ des distributions inititiales de X conduisant à une chaîne de Markov homogène pour l'agrégée Y, on va introduire quelques définitions et notations.

- Pour $i \in F$, on notera $\tilde{P}_i$ la matrice $(n(i), M)$ dont l'élément (j,k) est défini par: $\tilde{P}_i(j,k) = P(\, n(1) + \cdots + n(i-1) + j, B(k)\,)$, $1 \leq j \leq n(i)$, $k \in F$

Du Théorème 3.13, on déduit aisément la relation: $\hat{P}_j = (T_j.\pi^{B(j)})\tilde{P}_j$, où $\hat{P}_j$ désigne la $j^{ième}$ ligne de $\hat{P}$. Dans la suite, on définira toujours $\hat{P}$ de cette façon même si la chaîne agrégée n'est pas une chaîne de Markov homogène. C'est à dire que si Y est une chaîne de Markov homogène alors $\hat{P}$ est sa matrice des probabilités de transition et sinon, $\hat{P}_j \stackrel{\text{déf}}{=} (T_j.\pi^{B(j)})\tilde{P}_j$, $\forall j \in F$

- Pour $j \in F$, on notera σ_j le système linéaire $x_j\tilde{P}_j = \hat{P}_j$. Ce système a M équations et $n(j)$ inconnues. On vérifie que si x_j est une solution de σ_j alors $x_j\mathbb{1}^T = 1$. Par construction, pour tout $j \in F$, σ_j a au moins une solution qui est: $T_j.\pi^{B(j)}$.

- On regroupe les solutions de ces M systèmes dans l'ensemble suivant:

$$\mathcal{A}^* \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A}\,/\,T_l.\alpha^{B(l)} \text{ est solution de } \sigma_l \text{ pour tout } l \in F \text{ tel que } \alpha_{B(l)} \neq 0\,\}\,.$$

- On note aussi $\mathcal{A}_l^*$ l'ensemble des solutions positives du système σ_l.

Remarquons que $\mathcal{A}^* \neq \emptyset$ car par construction, $\pi \in \mathcal{A}^*$.

- On dira qu'un sous-ensemble $\mathcal{U}$ de $\mathcal{A}$ est *stable par* P si et seulement si $\forall x \in \mathcal{U}$ le vecteur $xP \in \mathcal{U}$.

Nous allons maintenant donner quelques propriétés de $\mathcal{A}^*$.

On rappelle qu'un polyèdre convexe de $\mathbb{R}^N$ est un sous-ensemble de $\mathbb{R}^N$ de la forme $\{x \in \mathbb{R}^N / \exists \mathrm{A}$, matrice $\mathrm{H} \times \mathrm{N}$, et $\mathrm{b} \in \mathbb{R}^N$ tel que $\mathrm{Ax}^T \leq \mathrm{b}^T\}$. Un polytope de $\mathbb{R}^N$ est un polyèdre convexe borné de $\mathbb{R}^N$ (voir [4] par exemple).

L'ensemble $\mathcal{A}_l^*$ des solutions positives du système σ_l est alors un polytope. Il s'en suit facilement le lemme suivant. Posons $T_l^{-1}.\mathcal{A}_l^* = \{T_l^{-1}.x / x \in \mathcal{A}_l^*\}$. Observons que l'ensemble $T_l^{-1}.\mathcal{A}_l^*$ est aussi un polytope.

Lemme 3.18 *L'ensemble $\mathcal{A}^*$ est l'enveloppe convexe formée à partir des ensembles $T_l^{-1}.\mathcal{A}_l^*$, $l \in F$. C'est à dire,*

$$\mathcal{A}^* = \sum_{l \in F} \lambda_l T_l^{-1}.\mathcal{A}_l^*$$

$$\text{où } \lambda_l \geq 0 \quad \text{et} \quad \sum_{l \in F} \lambda_l = 1$$

En particulier, $\mathcal{A}^*$ est aussi un polytope.

Démonstration. Soit x_1 (resp. x_2, ..., x_M) un élément de $\mathcal{A}_1^*$ (resp. de $\mathcal{A}_2^*$, ..., $\mathcal{A}_M^*$). Soit λ_1, ..., λ_M M réels positifs avec $\sum_k \lambda_k = 1$. Par définition de $\mathcal{A}^*$ il est clair que $x = \sum_{l \in F} \lambda_l T_l^{-1}.x_l \in \mathcal{A}^*$ car $\forall l \in F$, $x^{B(l)} = T_l^{-1}.x_l$.

Réciproquement, soit $x \in \mathcal{A}^*$. Le vecteur x s'écrit (de manière unique) par le Lemme 2.1:

$$x = \sum_{l \in F} \lambda_l x^{B(l)} \qquad \text{où } \lambda_l = x_{B(l)} 1^T$$

que l'on peut réécrire

$$x = \sum_{l \in F} \lambda_l T_l^{-1} T_l.x^{B(l)}$$

Le résultat est alors immédiat puisque les ensembles $\mathcal{A}_l^*$ sont des polytopes. □

Un premier résultat important concernant l'agrégation faible est alors donné par la théorème suivant:

Théorème 3.19 *$\mathcal{A}^*$ est stable par $P \Longleftrightarrow \mathcal{A}_M = \mathcal{A}^*$.*

Démonstration. $\mathcal{A}_M$ étant stable par P, l'implication de gauche à droite est évidente. Pour prouver la réciproque, remarquons tout d'abord que le Théorème 3.9 peut s'écrire de la façon suivante:

$Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène $\Longleftrightarrow \forall l \in F$, $T_l.\beta$ est solution de σ_l pour tout $\beta \in \mathcal{A}(\alpha, B(l))$.

Il s'en suit clairement que $\mathcal{A}_M \subseteq \mathcal{A}^*$.

Soit maintenant $\alpha \in \mathcal{A}^*$. Considérons une suite $(B(i), B(j), \ldots)$ possible pour α. On a:

$\alpha^{B(i)} \in \mathcal{A}^*$ puisque $(\alpha^{B(i)})^{B(i)} = \alpha^{B(i)}$
$\alpha^{B(i)} P \in \mathcal{A}^*$ par hypothèse
$(\alpha^{B(i)} P)^{B(j)} \in \mathcal{A}^*$ pour tout $i, j \in F$ par définition de $\mathcal{A}^*$
$(\alpha^{B(i)} P)^{B(j)} P \in \mathcal{A}^*$ pour tout $i, j \in F$ par hypothèse

En poursuivant par récurrence cet argument, on obtient:

$\mathcal{A}(\alpha, B(l)) \subseteq \mathcal{A}^*$ pour tout $l \in F$, ce qui est équivalent à $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène (Théorème 3.9). Donc $\alpha \in \mathcal{A}_M$. □

Remarque Dans [1], les auteurs pensent avoir caractérisé l'ensemble $\mathcal{A}_M$ par la condition "$\mathcal{A}^*$ stable par P" ce qui est erroné comme nous l'avons montré explicitement dans [6]. Néanmoins, nous avons repris leur démarche, qui permet de raisonner à l'aide de systèmes linéaires, pour aboutir au théorème précédent.

En reprenant l'exemple de la page 12, on a:

$$\pi = (7/16, 3/16, 6/16)$$

$$\tilde{P}_1 = \begin{pmatrix} 1/4 & 3/4 \end{pmatrix} \quad \tilde{P}_2 = \begin{pmatrix} 0 & 1 \\ 7/8 & 1/8 \end{pmatrix}$$

$$\hat{P}_1 = \begin{pmatrix} 1/4 & 3/4 \end{pmatrix} \quad \hat{P}_2 = \begin{pmatrix} 7/12 & 5/12 \end{pmatrix}$$

On obtient alors:

$$T_1^{-1}.\mathcal{A}_1^* = \{(1,0,0)\} \text{ et } T_2^{-1}.\mathcal{A}_2^* = \{(0,1/3,2/3)\}$$

D'où, $\mathcal{A}^* = \{(1-3a, a, 2a),\ a \in [0,1/3]\}$. On a vu précédemment que $\alpha_a P = \alpha_b$ avec $b = (3-4a)/12$. Donc, $\mathcal{A}^*$ est stable par P et on en déduit que $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^*$.

Pour donner une caractérisation de l'ensemble $\mathcal{A}_{\mathcal{M}}$, on introduit la notation suivante pour tout $j \geq 1$:

- $\mathcal{A}^j \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A}^* \ / \ \forall \beta = f(\alpha,\ B(i_1), \ldots, B(i_k)) \text{ avec } k \leq j,\ \beta \in \mathcal{A}^*\}$

On peut remarquer que cet ensemble s'écrit aussi:

$$\mathcal{A}^j \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A} \ / \ \forall \beta = f(\alpha,\ B(i_1), \ldots, B(i_k)) \text{ avec } k \leq j,\ \beta \in \mathcal{A}^*\}$$

Avec cette notation, on a:

$$\mathcal{A}^1 = \{\alpha \in \mathcal{A}^* \ / \ \forall m \in F \text{ tel que } \alpha_{B(m)} \neq 0,\ \alpha^{B(m)} \in \mathcal{A}^*\} = \mathcal{A}^*$$

Il est simple de voir que:

$$\forall j \geq 1,\ \mathcal{A}^{j+1} \subseteq \mathcal{A}^j$$

Le théorème suivant donne une expression de l'ensemble $\mathcal{A}_{\mathcal{M}}$ en fonction des $\mathcal{A}^j$.

Théorème 3.20

$$\mathcal{A}_{\mathcal{M}} = \bigcap_{j \geq 1} \mathcal{A}^j$$

Démonstration. De même que pour celle du Théorème 3.19, elle s'appuit sur l'équivalence suivante: $Y = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène $\Longleftrightarrow \forall l \in F,\ \forall \beta \in \mathcal{A}(\alpha, B(l)),\ T_l.\beta$ est solution de σ_l.

En d'autres termes:

$$\mathcal{A}_{\mathcal{M}} = \{\alpha \in \mathcal{A}^* \ / \ \forall l \in F,\ \mathcal{A}(\alpha, B(l)) \subseteq \mathcal{A}^*\}$$

On a donc: $\forall j \geq 1,\ \mathcal{A}_{\mathcal{M}} \subseteq \mathcal{A}^j$.

Réciproquement, si $\alpha \in \mathcal{A}^j\ \forall j \geq 1$, alors $\forall l \in F\ \mathcal{A}(\alpha, B(l)) \subseteq \mathcal{A}^*$; ce qui signifie que $\alpha \in \mathcal{A}_{\mathcal{M}}$ □

Une condition nécéssaire évidente pour avoir agrégation faible (i.e. pour que $\mathcal{A}_{\mathcal{M}} \neq \emptyset$) est donc que $\forall j \geq 1,\ \mathcal{A}^j \neq \emptyset$.

On va maintenant donner quelques propriétés des ensembles $\mathcal{A}^j$. On a déjà vu que $\mathcal{A}^1 = \mathcal{A}^*$ et que $\mathcal{A}^{j+1} \subseteq \mathcal{A}^j$, pour tout $j \geq 1$.

Le lemme suivant nous donne $\mathcal{A}^{j+1}$ en fonction de $\mathcal{A}^j$.

Lemme 3.21 *Pour tout $j \geq 1$:*

$$\mathcal{A}^{j+1} = \{\alpha \in \mathcal{A}^j / \alpha^{B(l)}P \in \mathcal{A}^j \text{ pour tout } l \in F \text{ tel que } \alpha_{B(l)} \neq 0\}$$

Démonstration. Soit $\alpha \in \mathcal{A}^{j+1}$. Soit $l \in F$ tel que $\alpha_{B(l)} \neq 0$ et posons $\beta = \alpha^{B(l)}P$. Soit $(B(i_1), \ldots, B(i_k))$ une suite possible pour β avec $k \leq j$.

$$f(\beta,\ B(i_1),\ \ldots,\ B(i_k)) = f(\alpha,\ B(l),\ B(i_1),\ \ldots,\ B(i_k)) \in \mathcal{A}^*$$

car $\alpha \in \mathcal{A}^{j+1}$ et $k \leq j$.

Réciproquement, soit $\alpha \in \mathcal{A}^j$ tel que $\alpha^{B(l)}P \in \mathcal{A}^j\ \forall l$ tel que $\alpha_{B(l)} \neq 0$. Soit $(B(l_1),\ \ldots,\ B(l_k))$ une suite possible pour α avec $k \leq j+1$.

$$f(\alpha,\ B(l_1),\ \ldots,\ B(l_k)) = f(\alpha^{B(l_1)}P,\ B(l_2),\ \ldots,\ B(l_k)) \in \mathcal{A}^*$$

car par hypothèse $\alpha^{B(l_1)}P \in \mathcal{A}^j$ et $k \leq j+1$. □

Comme corollaire de ce résultat, nous avons:

Lemme 3.22 *Pour tout $j \geq 1$, l'ensemble est $\mathcal{A}^j$ est convexe.*

Démonstration. Nous montrons la propriété par récurrence sur j. Dans le Lemme 3.18 on a prouvé la convexité de $\mathcal{A}^1$. Supposons donc la convexité de $\mathcal{A}^j$. Soit α et β deux éléments de $\mathcal{A}^{j+1}$ et soit λ et μ deux réels de $[0, 1]$ tels que $\lambda + \mu = 1$. Posons $\gamma = \lambda\alpha + \mu\beta$; par hypothèse, $\gamma \in \mathcal{A}^j$. Alors, pour tout $l \in F$ tel que $\gamma_{B(l)} \neq 0$ il est clair que $\gamma^{B(l)}P \in \mathcal{A}^j$ en appliquant le Lemme 2.2 ce qui donne $\gamma \in \mathcal{A}^{j+1}$.

Notons que si $\alpha^{B(l)} = 0$ (resp. $\beta^{B(l)} = 0$) alors $\gamma^{B(l)} = \beta^{B(l)}$ (resp. $\gamma^{B(l)} = \alpha^{B(l)}$) et le résultat suit immédiatement. $\square$

Une autre propriété des ensembles $\mathcal{A}^j$ est donnée par le lemme suivant.

Lemme 3.23 *Pour tout $j \geq 1$, on a:*

$$\alpha \in \mathcal{A}^j \Longleftrightarrow \alpha^{B(l)} \in \mathcal{A}^j \text{ pour tout } l \in F \text{ tel que } \alpha_{B(l)} \neq 0$$

Démonstration. Soit $\alpha \in \mathcal{A}^j$ et $l \in F$ tel que $\alpha_{B(l)} \neq 0$. Pour tout $k \leq j$ et pour toute suite $(B(l),\ B(i_2),\ \ldots,\ B(i_k))$ possible pour α, on a:

$$f(\alpha^{B(l)},\ B(l),\ B(i_2),\ \ldots,\ B(i_k)) = f(\alpha,\ B(l),\ B(i_2),\ \ldots,\ B(i_k)) \in \mathcal{A}^1$$

car $k \leq j$ et $\alpha \in \mathcal{A}^j$. Donc $\alpha^{B(l)} \in \mathcal{A}^j$.

La réciproque est évidente étant donné que tout élément $\alpha \in \mathcal{A}$ s'écrit de manière unique (Lemme 2.1)

$$\alpha = \sum_{l \in F} \alpha_{B(l)} 1^T \alpha^{B(l)}$$

et que $\mathcal{A}^j$ est convexe. $\square$

Une condition suffisante pour avoir agrégation faible (i.e. pour que $\mathcal{A}_{\mathcal{M}} \neq \emptyset$) est donnée par le corollaire (du Théorème 3.20) suivant.

Corollaire 3.24 *Si $\exists j \geq 1$ tel que $\mathcal{A}^{j+1} = \mathcal{A}^j$ alors $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^{j+k}$, $\forall k \geq 0$.*

Démonstration. Si $\mathcal{A}^j = \emptyset$, le résultat est trivial.

Soit $\mathcal{U} = \mathcal{A}^j = \mathcal{A}^{j+1}$ et $\alpha \in \mathcal{U}$. Par définition, pour tout $i_1, \ldots, i_{j+1} \in F$, $f(\alpha,\ B(i_1), \ldots, B(i_{j+1})) \in \mathcal{A}^*$.

On a alors:

$$\alpha \in \mathcal{U} \Longrightarrow \alpha^{B(l)} P \in \mathcal{U},\ \forall l \in F \text{ tel que } \alpha_{B(l)} \neq 0 \quad \text{par le Lemme 3.23}$$

Cela signifie que si $\alpha \in \mathcal{U}$ alors $\forall l \in F$, $\mathcal{A}(\alpha, B(l)) \subseteq \mathcal{U}$, ce qui implique que $\alpha \in \mathcal{A}_{\mathcal{M}}$. Enfin, grâce au Théorème 3.20, on obtient le résultat. $\square$

On peut maintenant donner une généralisation du Théorème 3.20.

Théorème 3.25 *$\mathcal{A}^j$ est stable par $P \Longleftrightarrow \mathcal{A}_{\mathcal{M}} = \mathcal{A}^j$*

Démonstration. La démonstration est identique à celle du Théorème 3.20.
□

Introduisons les notations suivantes: pour tout $j \geq 1$ et pour tout $l \in F$,

- $\mathcal{A}_l^j = \{T_l.\alpha^{B(l)} / \alpha \in \mathcal{A}^j \text{ et } \alpha_{B(l)} \neq 0\}$
- $T_l^{-1}.\mathcal{A}_l^j = \{T_l^{-1}.x / x \in \mathcal{A}_l^j\}$

On en déduit que: $\forall j \geq 1, \forall l \in F$,

- $\mathcal{A}_l^* = \mathcal{A}_l^1$
- $\mathcal{A}_l^{j+1} \subseteq \mathcal{A}_l^j$
- $T_l^{-1}.\mathcal{A}_l^j = \{\alpha^{B(l)} / \alpha \in \mathcal{A}^j \text{ et } \alpha_{B(l)} \neq 0\} \subseteq \mathcal{A}^j$

Avec ces notations, on a:

Lemme 3.26 *Si $l \in F$ et $j \geq 1$, alors:*

$$\mathcal{A}_l^{j+1} = \{x \in \mathcal{A}_l^j / (T_l^{-1}.x)P \in \mathcal{A}^j\}$$

Démonstration. Soit $x \in \mathcal{A}_l^{j+1}$. Par définition, il existe un $\alpha \in \mathcal{A}^{j+1}$, $\alpha_{B(l)} \neq 0$ tel que $x = T_l.\alpha^{B(l)}$. On a donc $x \in \mathcal{A}_l^j$ et $(T_l^{-1}.x)P = [T_l^{-1}.(T_l.\alpha^{B(l)})]P = \alpha^{B(l)}P$. Soit $k \leq j$ et $(B(i_1), \ldots, B(i_k))$ une suite possible pour $\alpha^{B(l)}P$, on a:

$$f(\alpha^{B(l)}P, B(i_1), \ldots, B(i_k)) = f(\alpha, B(l), B(i_1), \ldots, B(i_k)) \in \mathcal{A}^1$$

car $k \leq j$ et $\alpha \in \mathcal{A}^{j+1}$. Donc, $\alpha^{B(l)}P \in \mathcal{A}^j$.

Réciproquement, soit $x \in \mathcal{A}_l^j$ tel que $(T_l^{-1}.x)P \in \mathcal{A}^j$. Le vecteur $x \in \mathcal{A}_l^j$ donc il existe $\alpha \in \mathcal{A}^j$, $\alpha_{B(l)} \neq 0$ tel que $x = T_l.\alpha^{B(l)}$. Par le lemme 3.23, on peut prendre $\alpha = T_l^{-1}.x$. On obtient alors $(T_l^{-1}.x)P = \alpha P = \alpha^{B(l)}P \in \mathcal{A}^j$ ce qui signifie que $\alpha \in \mathcal{A}^{j+1}$ par le Lemme 3.21 et donc que $x \in \mathcal{A}_l^{j+1}$ □

Lemme 3.27 *Pour tout $j \geq 1$, l'ensemble $\mathcal{A}^j$ est l'enveloppe convexe formée à partir des ensembles $T_l^{-1}.\mathcal{A}_l^j$, $l \in F$. C'est à dire,*

$$\mathcal{A}^j = \sum_{l \in F} \lambda_l T_l^{-1}.\mathcal{A}_l^j$$

$$\textit{où } \lambda_l \geq 0 \quad \textit{et} \quad \sum_{l \in F} \lambda_l = 1$$

Démonstration. Soit x_1 (resp. $x_2, \ldots, x_M$) un élément de $\mathcal{A}_1^j$ (resp. de $\mathcal{A}_2^j, \ldots, \mathcal{A}_M^j$). Soit $\lambda_1, \ldots, \lambda_M$ M réels positifs avec $\sum_k \lambda_k = 1$. On a vu que $T_l^{-1}.\mathcal{A}_l^j \subseteq \mathcal{A}^j$ pour tout $l \in F$ et que l'ensemble $\mathcal{A}^j$ est un convexe (Lemme 3.22) donc

$$x = \sum_{l \in F} \lambda_l T_l^{-1}.x_l \in \mathcal{A}^j$$

Réciproquement, soit $x \in \mathcal{A}^j$. Le vecteur x s'écrit (de manière unique) par le Lemme 2.1:

$$x = \sum_{l \in F} \lambda_l x^{B(l)} \qquad \text{où } \lambda_l = x_{B(l)} 1^T$$

Le Lemme 3.23 permet alors de conclure. □

Lemme 3.28 *Pour tout $j \geq 1$, pour tout $l \in F$, $\mathcal{A}_l^j$ et $\mathcal{A}^j$ sont des polytopes.*

Démonstration. On démontre ce lemme par récurrence. On a vu précédemment (Lemme 3.18) que pour $j = 1$, le lemme est vrai. On a vu aussi (Lemme 3.26) que $\mathcal{A}_l^{j+1} = \{x \in \mathcal{A}_l^j / (T_l^{-1}.x)P \in \mathcal{A}^j\}$. Supposons le lemme vrai jusquà l'ordre j. Le Lemme 3.27 nous permet alors d'affirmer que $\mathcal{A}^j$ est un polytope. Soit $l \in F$ et Φ_l l'application suivante:

$$\begin{array}{rccc} \Phi_l : & \mathbb{R}^{n(l)} & \longrightarrow & \mathbb{R}^N \\ & x & \longmapsto & (T_l^{-1}.x)P \end{array}$$

L'application Φ_l est une application linéaire et donc l'ensemble $\Phi_l^{-1}(\mathcal{A}^j) = \{x \in \mathbb{R}^{n(l)} / (T_l^{-1}.x)P \in \mathcal{A}^j\}$ est un polytope (voir [4], pages 18 et 174). Finalement, $\mathcal{A}_l^{j+1} = \mathcal{A}_l^j \cap \Phi_l^{-1}(\mathcal{A}^j)$ est aussi un polytope par l'hypothèse de récurrence.

On en déduit de façon immédiate (Lemme 3.27) que pour tout $j \geq 1$, l'ensemble $\mathcal{A}^j$ est un polytope. □

Nous donnons maintenant la caractérisation finale de l'ensemble $\mathcal{A}_M$. Cette caractérisation a la forme d'un algorithme qui permet d'obtenir $\mathcal{A}_M$ au bout d'un nombre fini de pas. Plus précisement, on montrera que $\mathcal{A}_M = \mathcal{A}^N$ où, rappelons-le, N est le nombre d'états de la chaîne X. Pour cela nous aurons besoin des lemmes suivants.

Lemme 3.29 *Soit α un point de $\mathcal{A}^{j+1}$ et β un point de $\mathcal{A}^j$ n'appartenant pas à $\mathcal{A}^{j+1}$, alors aucun point du segment de droite $]\alpha\ , \beta]$ n'appartient à $\mathcal{A}^{j+1}$.*

Démonstration. On démontre ce lemme par l'absurde. Soit γ un point du segment $]\alpha\ ,\ \beta]$, c'est à dire, il existe $\lambda \in [0,1[$ tel que:

$$\gamma = \lambda\alpha + (1-\lambda)\beta$$

Supposons que $\gamma \in \mathcal{A}^{j+1}$. On va montrer alors que $\beta \in \mathcal{A}^{j+1}$ ce qui est contraire à l'hypothèse. Pour montrer que $\beta \in \mathcal{A}^{j+1}$, il suffit de montrer (Lemme 3.21) que $\forall l \in F$ tel que $\beta_{B(l)} \neq 0$, $\beta^{B(l)}P \in \mathcal{A}^j$; c'est à dire $\forall l \in F$ tel que $\beta_{B(l)} \neq 0$, $\forall\sigma$ suite possible pour $\beta^{B(l)}P$ de longueur inférieure ou égale à j, $f(\beta^{B(l)}P, \sigma) \in \mathcal{A}^1$.

Soit donc $l \in F$ tel que $\beta_{B(l)} \neq 0$. Cela implique en particulier que $\gamma_{B(l)} \neq 0$. Soit σ une suite possible pour $\beta^{B(l)}P$ de longueur inférieure ou égale à j.

Si $\alpha_{B(l)} = 0$ alors $\gamma^{B(l)} = \beta^{B(l)}$; donc $f(\beta^{B(l)}P, \sigma) = f(\gamma^{B(l)}P, \sigma) \in \mathcal{A}^1$.

Si $\alpha_{B(l)} \neq 0$ et si σ n'est pas possible pour $\alpha^{B(l)}P$ alors par le Lemme 3.4, on a: $f(\beta^{B(l)}P, \sigma) = f(\gamma^{B(l)}P, \sigma) \in \mathcal{A}^1$.

Enfin si $\alpha_{B(l)} \neq 0$ et si σ est possible pour $\alpha^{B(l)}P$ alors σ est aussi possible pour $\gamma^{B(l)}P$ et donc $f(\gamma^{B(l)}P, \sigma) \in \mathcal{A}^1$. En appliquant maintenant le Lemme 3.5 on a:

$$\begin{aligned}
f(\gamma^{B(l)}P, \sigma) &= f(\gamma,\ B(l),\ \sigma) \\
&= \lambda\frac{K_\alpha}{K}f(\alpha,\ B(l),\ \sigma) + (1-\lambda)\frac{K_\beta}{K}f(\beta,\ B(l),\ \sigma) \\
&= \lambda\frac{K_\alpha}{K}f(\alpha^{B(l)}P, \sigma) + (1-\lambda)\frac{K_\beta}{K}f(\beta^{B(l)}P, \sigma)
\end{aligned}$$

où K_α et K_β sont donnés dans le Lemme 3.5 et $K = \lambda K_\alpha + (1-\lambda)K_\beta$. Mais, $f(\gamma^{B(l)}P, \sigma) \in \mathcal{A}^1$ et $f(\alpha,\ B(l),\ \sigma) \in \mathcal{A}^1$. Si l'on note $B(m)$ le dernier élément de la suite σ, alors on a:

$$\begin{aligned}
(1-\lambda)\frac{K_\beta}{K}T_m.f(\beta^{B(l)}P, \sigma)\tilde{P}_m &= T_m.f(\gamma^{B(l)}P, \sigma)\tilde{P}_m - \lambda\frac{K_\alpha}{K}T_m.f(\alpha^{B(l)}P, \sigma)\tilde{P}_m \\
&= \hat{P}_m - \lambda\hat{P}_m \\
&= (1-\lambda)\hat{P}_m
\end{aligned}$$

λ étant $\neq 1$, on obtient $T_m.f(\beta^{B(l)}P, \sigma)\tilde{P}_m = \hat{P}_m$ ce qui signifie que $f(\beta^{B(l)}P, \sigma) \in \mathcal{A}^1$.

□

Remarque Il est simple de voir que ce lemme reste vrai pour $j = 0$ si l'on pose $\mathcal{A}^0 \stackrel{\text{déf}}{=} \mathcal{A}$.

Lemme 3.30 *Si α et β sont deux points distincts de $\mathcal{A}^{j+1}$ alors*

$$\left[D(\alpha,\beta)\cap \mathcal{A}^j\right]\subseteq \mathcal{A}^{j+1}$$

où $D(\alpha,\beta)$ est la droite passant par les points α et β.

Démonstration. On sait déjà que $\mathcal{A}^{j+1}$ est convexe, donc $[\alpha,\beta]\subseteq \mathcal{A}^{j+1}$. Supposons qu'il existe un point $\gamma\notin[\alpha,\beta]$ tel que $\gamma\in D(\alpha,\beta)\cap\mathcal{A}^j$ et $\gamma\notin\mathcal{A}^{j+1}$. Supposons que le point β se trouve entre les points α et γ. On a alors par le lemme précédent: aucun point du segment de droite $]\alpha,\gamma]$ n'appartient à $\mathcal{A}^{j+1}$. Or, par hypothèse, β appartient à $\mathcal{A}^{j+1}$. □

On appelle $intrel(C)$ l'intérieur de C relativement au plus petit sous-epace affine S le contenant et de la même manière $frontrel(C)$ est la frontière de C relativement à S. On a alors:

Lemme 3.31 *Pour tout $j\geq 1$, si $\mathcal{A}^{j+1}$ n'est pas réduit à un seul élément alors $frontrel(\mathcal{A}^{j+1})\subseteq frontrel(\mathcal{A}^j)$.*

Démonstration. La proposition est triviale si $\mathcal{A}^{j+1}=\mathcal{A}^j$ ou si $\mathcal{A}^{j+1}=\emptyset$. Soit donc $\emptyset\neq\mathcal{A}^{j+1}\subset\mathcal{A}^j$. Soit $\alpha\in frontrel(\mathcal{A}^{j+1})$ et supposons que $\alpha\notin frontrel(\mathcal{A}^j)$. Par hypothèse, il existe un point $\beta\in intrel(\mathcal{A}^{j+1})$. En appliquant le lemme précédent, on en déduit que $D(\alpha,\beta)\cap\mathcal{A}^j\subseteq\mathcal{A}^{j+1}$; donc $D(\alpha,\beta)\cap\mathcal{A}^j$ est un segment de droite $[a,b]$ avec $a,b\in\left(frontrel(\mathcal{A}^j)\cap frontrel(\mathcal{A}^{j+1})\right)$, d'où $\alpha\in]a,b[$, ce qui contredit l'hypothèse. □

Lemme 3.32 *Soit α un point de $\mathcal{A}^{j+1}$ et β un point de $\mathcal{A}^j$ n'appartenant pas à $\mathcal{A}^{j+1}$ alors aucun point de la droite $D(\alpha,\beta)$ autre que α n'appartient à $\mathcal{A}^{j+1}$.*

Démonstration. Si un point γ de la droite $D(\alpha,\beta)$ autre que α était dans $\mathcal{A}^{j+1}$, alors par le Lemme 3.30 on aurait $\beta\in\mathcal{A}^{j+1}$ ce qui contredit l'hypothèse. □

La dimension d'un convexe C de $\mathbb{R}^N$ est la dimension du plus petit sous-espace affine le contenant; on la notera $dim(C)$. Nous avons alors le lemme suivant qui nous assurera que le processus de détermination de $\mathcal{A}_M$ est fini.

Lemme 3.33 *Soit $j\geq 1$ tel que $\mathcal{A}^{j+1}\neq\emptyset$ et $\mathcal{A}^{j+1}\neq\mathcal{A}^j$; alors*

$$dim(\mathcal{A}^{j+1})<dim(\mathcal{A}^j)$$

Démonstration.

$$dim(\mathcal{A}^{j+1}) \leq dim(\mathcal{A}^{j}) \quad \text{car } \mathcal{A}^{j+1} \subseteq \mathcal{A}^{j}$$

Supposons que l'on ait égalité entre ces deux dimensions. Cette dimension commune ne peut pas être nulle par hypothèse et par les propriétés des ensembles $\mathcal{A}^j$. Il existe donc un point $x \in intrel(\mathcal{A}^{j+1})$. Soit y un point de $\mathcal{A}^j - \mathcal{A}^{j+1}$. Etant donné que $dim(\mathcal{A}^{j+1}) = dim(\mathcal{A}^j)$, il existe $z \in intrel(\mathcal{A}^{j+1})$ tel que $z \neq x$ et $z \in D(x,y)$. Par le Lemme 3.30,

$$\left[D(x,z) \cap \mathcal{A}^j\right] \subseteq \mathcal{A}^{j+1}$$

et ceci contredit le fait que $y \notin \mathcal{A}^{j+1}$. □

Remarque Les résultats précédents sont aussi valables si l'on considère les sous-ensembles $\mathcal{A}_l^j$ plutôt que $\mathcal{A}^j$. Les énoncés correspondants sont (sans les démonstrations):

- Soit x un point de $\mathcal{A}_l^{j+1}$ et y un point de $\mathcal{A}_l^j$ n'appartenant pas à $\mathcal{A}_l^{j+1}$ alors aucun point de la droite $D(x,y)$ autre que x n'appartient à $\mathcal{A}_l^{j+1}$.
- Si x et y sont deux points distincts de $\mathcal{A}_l^{j+1}$ alors

$$\left[D(x,y) \cap \mathcal{A}_l^j\right] \subseteq \mathcal{A}_l^{j+1}$$

- Si $\mathcal{A}_l^{j+1}$ n'est pas réduit à un seul point alors $frontrel(\mathcal{A}_l^{j+1}) \subseteq frontrel(\mathcal{A}_l^j)$.
- Si $\mathcal{A}_l^{j+1} \neq \emptyset$ et $\mathcal{A}_l^{j+1} \neq \mathcal{A}_l^j$ alors $dim(\mathcal{A}_l^{j+1}) < dim(\mathcal{A}_l^j)$.

Nous pouvons maintenant prouver le théorème annoncé.

Théorème 3.34

$$\mathcal{A}_{\mathcal{M}} = \mathcal{A}^N$$

Démonstration. Si $\mathcal{A}^1 = \mathcal{A}$ le résultat est immédiat: $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^j$ pour tout j. Sinon, observons la suite $\mathcal{A}^1, \ldots, \mathcal{A}^N$. Si deux de ses éléments sont égaux, alors $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^N$ par le Corollaire 3.24. Si tous ses éléments sont distincts, alors, par le résultat précédent, les dimensions décroissent strictement. Par la remarque suivant le Lemme 3.29, on peut aisément vérifier que $dim(\mathcal{A}^1) < dim(\mathcal{A}) = N - 1$. On a donc: $dim(\mathcal{A}^j) < N - j$ pour $j = 1, 2, \ldots, N - 1$, d'où $\mathcal{A}^N = \emptyset$. Le Théorème 3.20 permet alors de conclure. □

Algorithme de calcul de $\mathcal{A}_{\mathcal{M}}$.

Pour déterminer l'ensemble $\mathcal{A}_{\mathcal{M}}$, on procède comme suit. Supposons que la chaîne X ne soit pas fortement agrégeable par rapport à la partition $\mathcal{B}$. La première étape consiste à calculer l'ensemble $\mathcal{A}^1$ et à déterminer s'il est stable par P. Dans l'affirmative, le Théorème 3.19 entraîne $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^1$. Si $\mathcal{A}^1$ n'est pas stable par P, on vérifie d'abord si $\forall l \in F$ le vecteur $\pi^{B(l)}P$ est dans $\mathcal{A}^1$. Si $\exists l \in F$ tel que $\pi^{B(l)}P \notin \mathcal{A}^1$, alors $\pi \notin \mathcal{A}^2$ et $\mathcal{A}_{\mathcal{M}} = \emptyset$ (conséquence des Théorèmes 3.13 et 3.20). Si $\pi \in \mathcal{A}^2$ alors on calcule cet ensemble et on répète les opérations précédentes. On peut exprimer ceci sous forme "algorithmique" de la façon suivante. Notons Ψ la procédure définie par:

$$\begin{array}{rrcl} \Psi : & \mathcal{P}(\mathcal{A}) & \longrightarrow & \mathcal{P}(\mathcal{A}) \\ & \mathcal{U} \subseteq \mathcal{A} & \longmapsto & \left\{\alpha \in \mathcal{U} / \forall l \in F \text{ tel que } \alpha_{B(l)} \neq 0 \text{ on a } \alpha^{B(l)}P \in \mathcal{U}\right\} \end{array}$$

Avec cette notation, $\mathcal{A}^{j+1} = \Psi(\mathcal{A}^j)$ (Lemme 3.21). L'algorithme peut alors s'écrire:

si X est fortement agrégeable **alors** $\mathcal{A}_{\mathcal{M}} = \mathcal{A}$; **arrêt**
sinon $\mathcal{U} := \mathcal{A}^1$;
    boucle
        si $\mathcal{U}$ est stable par P **alors sortirpar** *nonvide* **finsi** ;
        si $\exists l \in F$: $\pi^{B(l)}P \notin \mathcal{U}$ **alors sortirpar** *vide* **finsi** ;
        $\mathcal{U} := \Psi(\mathcal{U})$
    finboucle
    sorties
        sortie *nonvide* : $\mathcal{A}_{\mathcal{M}} := \mathcal{U}$
        sortie *vide* : $\mathcal{A}_{\mathcal{M}} := \emptyset$
    finsorties
finsi

Le Théorème 3.34 nous assure que le nombre d'itérations est borné par N.

Illustrons ceci par un exemple. Soit la famille de chaînes de Markov homogènes $(\,.\,, P)$ avec:

$$P = \begin{pmatrix} 3/14 & 3/14 & 3/14 & 3/14 & 1/7 \\ 1/6 & 1/6 & 1/6 & 1/3 & 1/6 \\ 1/8 & 3/8 & 1/4 & 1/6 & 1/12 \\ 3/8 & 1/8 & 1/4 & 1/6 & 1/12 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$$

On a:

$$\pi = \frac{1}{193}(42, 42, 42, 42, 25)$$

Considérons la partition $\mathcal{B} = \{B(1), B(2)\}$, où $B(1) = \{1,2,3,4\}$ et $B(2) = \{5\}$.

On a alors:

$$\tilde{P}_1 = \begin{pmatrix} 6/7 & 1/7 \\ 5/6 & 1/6 \\ 11/12 & 1/12 \\ 11/12 & 1/12 \end{pmatrix} \qquad \tilde{P}_2 = \begin{pmatrix} 4/5 & 1/5 \end{pmatrix}$$

$$\pi^{B(1)} = (1/4, 1/4, 1/4, 1/4, 0) \qquad \pi^{B(2)} = (0,0,0,0,1)$$

On en déduit:

$$\hat{P}_1 = \begin{pmatrix} 37/42 & 5/42 \end{pmatrix} \qquad \hat{P}_2 = \begin{pmatrix} 4/5 & 1/5 \end{pmatrix}$$

On vérifie d'abord que $\mathcal{A}_\pi$ stable par P, d'où $\mathcal{A}_\pi \subseteq \mathcal{A}_\mathcal{M}$, c'est à dire $\mathcal{A}_\pi \subseteq \mathcal{A}^j$, $\forall j \geq 1$

On calcule ensuite $\mathcal{A}^1$. Les sommets de $\mathcal{A}^1$ sont les points suivants:

- $\alpha_1 = (0, 3/7, 0, 4/7, 0)$
- $\alpha_2 = (3/5, 0, 0, 2/5, 0)$
- $\alpha_3 = (3/5, 0, 2/5, 0, 0)$
- $\alpha_4 = (0, 3/7, 4/7, 0, 0)$
- $\alpha_5 = (0, 0, 0, 0, 1)$

On vérifie alors que: $\alpha_2, \alpha_5 \in \mathcal{A}^2$ et $\alpha_1, \alpha_3, \alpha_4 \notin \mathcal{A}^2$. On sait de plus que le point $\pi^{B(1)} \in \mathcal{A}^2$ et que $\pi^{B(1)}$ ne se trouve pas sur la droite $D(\alpha_2, \alpha_5)$; donc nécéssairement $dim(\mathcal{A}^2) = 2$. Le troisième sommet α_6 de $\mathcal{A}^2$ se trouve donc sur la droite $D(\alpha_2, \pi^{B(1)})$. On peut vérifier que $\alpha_6 = (0, 3/7, 3/7, 1/7, 0)$.

Les sommets de $\mathcal{A}^2$ sont donc $\alpha_2, \alpha_5, \alpha_6$. On peut vérifier alors que les points α_2 et α_6 n'appartiennent pas à $\mathcal{A}^3$ et que α_5 appartient à $\mathcal{A}^3$. Donc, $\mathcal{A}^3 = \mathcal{A}_\pi = \mathcal{A}_\mathcal{M}$ qui est le segment $[\pi^{B(1)}, \pi^{B(2)}]$.

4 Le cas continu

$X = (X_t)_{t \geq 0}$ désigne un processus markovien homogène sur l'espace d'état E. Le processus X est maintenant donné par sa matrice des taux de transitions (ou générateur infinitésimal) A. On pose $A(i,i) \stackrel{\text{déf}}{=} -\sum_{j \neq i} A(i,j)$. De la même manière que dans le cas discret, on notera (α, A) ce processus markovien homogène lorsque ce sera nécéssaire, α désignant la distribution de probabilité initiale de X. Par exemple, $P_\alpha(X_t \in B)$ désigne la probabilité que le processus markovien (α, A) soit dans le sous-ensemble B de E à l'instant t. La quantité $A(i, B)$ représente le taux de transition de l'état i vers le sous-ensemble B de E, c'est à dire: $A(i, B) = \sum_{j \in B} A(i, j)$. La quantité $e^{At}(i, B)$ représente la probabilité que le processus soit dans le sous-ensemble B de E à l'instant t sachant qu'il démarre dans l'état i, c'est à dire: $e^{At}(i, B) = \sum_{j \in B} e^{At}(i, j)$.

4.1 Préliminaires

Les lemmes préliminaires de la section précédente se traduisent de la façon suivante.

Lemme 4.1

$$P_\alpha(X_t \in B) \;=\; P_{\alpha e^{At}}(X_0 \in B)$$

Démonstration. Remarquons tout d'abord que $P_\beta(X_0 \in B) \;=\; \sum_{i \in B} \beta(i)$. Ainsi:

$$\begin{aligned} P_\alpha(X_t \in B) &= \sum_{i \in B} \pi_t(i) \quad \text{où } \pi_t = \alpha e^{At} \text{ représente} \\ & \qquad\qquad\qquad \text{la distribution de } X_t \\ &= P_{\pi_t}(X_0 \in B) \end{aligned}$$

□

Lemme 4.2

$$P_\alpha(X_t \in B / X_0 \in D) \;=\; P_{\alpha^D}(X_t \in B)$$

Démonstration.

$$\begin{aligned} P_\alpha(X_t \in B / X_0 \in D) &= \frac{\sum_{d \in D} \alpha(d) e^{At}(d, B)}{\sum_{d \in D} \alpha(d)} \\ &= \sum_{d \in D} \left(\frac{\alpha(d)}{\sum_{d' \in D} \alpha(d')} \right) e^{At}(d, B) \\ &= P_{\alpha^D}(X_t \in B) \end{aligned}$$

□

Une généralisation de ces relations est donnée par le lemme suivant.

Lemme 4.3

$$P_\alpha(X_{t+s} \in B/X_t \in D) \;=\; P_{\alpha e^{At}}(X_s \in B/X_0 \in D)$$

Démonstration.

$$\begin{aligned}
P_\alpha(X_{t+s} \in B/X_t \in D) &= \frac{\sum_{d\in D} P_\alpha(X_t = d)e^{As}(d,B)}{P_\alpha(X_t \in D)} \\
&= \sum_{d\in D}\left(\frac{P_{\alpha e^{At}}(X_0 = d)}{P_{\alpha e^{At}}(X_0 \in D)}\right) e^{As}(d,B) \quad \text{par le Lemme 4.1} \\
&= \sum_{d\in D} P_{(\alpha e^{At})^D}(X_0 = d)e^{As}(d,B) \\
&= P_{(\alpha e^{At})^D}(X_s \in B) \\
&= P_{\alpha e^{At}}(X_s \in B/X_0 \in D) \quad \text{par le Lemme 4.2}
\end{aligned}$$

□

Définition 4.4 *Une suite $(C_0, C_1, \ldots, C_j)$ de sous-ensembles de E sera dite possible pour la distribution initiale α si pour tout $0 = t_0 < t_1 < \cdots < t_j$, $P_\alpha(X_{t_0} \in C_0, X_{t_1} \in C_1, \ldots, X_{t_j} \in C_j) > 0$. En particulier, si $B \in \mathcal{B}$, (B) est une suite possible pour α si $\alpha_B \neq 0$.*

Définition 4.5 *Soit $\alpha \in \mathcal{A}$ et $(C_0, C_1, \ldots, C_j)$ une suite d'éléments de $\mathcal{B}$ possible pour α, on définit, pour tout $0 = t_0 < t_1 < \cdots < t_j$, le vecteur de probabilité $g(\alpha, C_0, (C_1, t_1), \ldots, (C_j, t_j))$ récursivement par:*

$$\begin{aligned}
g(\alpha, C) &= \alpha^C \\
g(\alpha, C_0, (C_1, t_1)) &= \left(g(\alpha, C_0)e^{At_1}\right)^{C_1} \\
&\textit{et pour } k \geq 2 : \\
g(\alpha, C_0, (C_1, t_1), \ldots, (C_k, t_k)) &= \left(g(\alpha, C_0, (C_1, t_1), \ldots, (C_{k-1}, t_{k-1}))e^{A(t_k - t_{k-1})}\right)^{C_k}
\end{aligned}$$

Par exemple, $g(\alpha, B, (C, t_1), (D, t_2)) = \left((\alpha^B e^{At_1})^C e^{A(t_2 - t_1)}\right)^D$.

Lemme 4.6 *La restriction de $\beta = g(\alpha, C_0, (C_1, t_1), \ldots, (C_j, t_j))$ à C_j est donnée par l'expression suivante:*

$$\beta_{C_j} = K^{-1}\alpha_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2 - t_1)}_{C_1C_2} \cdots e^{A(t_j - t_{j-1})}_{C_{j-1}C_j}$$

$$\text{où } K = \alpha_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2 - t_1)}_{C_1C_2} \cdots e^{A(t_j - t_{j-1})}_{C_{j-1}C_j} 1^T$$

Démonstration. Pour $j = 0$, aucune difficulté. Supposons que la relation soit vraie pour toute suite formée de n sous-ensembles et notons:

$$\gamma = g(\alpha, C_0, (C_1, t_1), \ldots, (C_{n-1}, t_{n-1}))$$

On a alors:

$$\gamma_{C_{n-1}} = L^{-1} \alpha_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2-t_1)}_{C_1C_2} \ldots e^{A(t_{n-1}-t_{n-2})}_{C_{n-2}C_{n-1}}$$
$$\text{où } L = \alpha_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2-t_1)}_{C_1C_2} \ldots e^{A(t_{n-1}-t_{n-2})}_{C_{n-2}C_{n-1}} 1^T.$$

Si $\beta = g(\alpha, C_0, (C_1, t_1), \ldots, (C_n, t_n)) = (\gamma e^{A(t_n-t_{n-1})})^{C_n}$, on a:

$$\beta_{C_n} = H^{-1}(\gamma e^{A(t_n-t_{n-1})})_{C_n} \quad \text{avec } H = (\gamma e^{A(t_n-t_{n-1})})_{C_n} 1^T$$

Mais $(\gamma e^{A(t_n-t_{n-1})})_{C_n} = \gamma_{C_{n-1}} e^{A(t_n-t_{n-1})}_{C_{n-1}C_n}$ puisque $\gamma(i) = 0$ si $i \notin C_{n-1}$. Donc,

$$\begin{aligned} \beta_{C_n} &= H^{-1} \gamma_{C_{n-1}} e^{A(t_n-t_{n-1})}_{C_{n-1}C_n} \\ &= H^{-1} L^{-1} \alpha_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2-t_1)}_{C_1C_2} \ldots e^{A(t_{n-1}-t_{n-2})}_{C_{n-2}C_{n-1}} e^{A(t_n-t_{n-1})}_{C_{n-1}C_n} \end{aligned}$$

D'où le résultat puisque K^{-1} est simplement une constante de normalisation.
□

Comme dans le cas discret, on a la propriété de convexité suivante.

Lemme 4.7 *Soit $(C_0, \ldots, C_j)$ une suite d'éléments de $\mathcal{B}$. Soit $\mathcal{A}(C_0, \ldots, C_j) \stackrel{\text{déf}}{=} \{\gamma \in \mathcal{A} /$ la suite $(C_0, \ldots, C_j)$ est possible pour $\gamma\}$.*

$\mathcal{A}(C_0, \ldots, C_j)$ est convexe et s'il est non vide, on a la propriété suivante: pour tout $0 < t_1 < \ldots < t_j$, pour tout $\alpha, \beta \in \mathcal{A}(C_0, \ldots, C_j)$ et pour tout $\lambda, \mu \geq 0$ tels que $\lambda + \mu = 1$,

$$g(\lambda\alpha+\mu\beta, C_0, \ldots, (C_j, t_j)) = \lambda \frac{K_\alpha}{K} g(\alpha, C_0, \ldots, (C_j, t_j)) + \mu \frac{K_\beta}{K} g(\beta, C_0, \ldots, (C_j, t_j))$$

$$\text{où } K_\alpha = \alpha_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2-t_1)}_{C_1C_2} \ldots e^{A(t_{j-1}-t_{j-2})}_{C_{j-2}C_{j-1}} e^{A(t_j-t_{j-1})}_{C_{j-1}C_j} 1^T$$
$$K_\beta = \beta_{C_0} e^{At_1}_{C_0C_1} e^{A(t_2-t_1)}_{C_1C_2} \ldots e^{A(t_{j-1}-t_{j-2})}_{C_{j-2}C_{j-1}} e^{A(t_j-t_{j-1})}_{C_{j-1}C_j} 1^T$$
$$\text{et } K = \lambda K_\alpha + \mu K_\beta.$$

Démonstration. La démonstration est identique à celle du Lemme 3.5. □

On va donner maintenant une généralisation du Lemme 4.3.

Lemme 4.8 *Pour tout $k \geq 0$, on a:*

$$P_\alpha(X_{t_{n+1}} \in B / X_{t_n} \in C_n, \ldots, X_{t_{n-k}} \in C_{n-k}) = P_\beta(X_{t_{n+1}-t_n} \in B) \quad \textit{pour tout } n \geq k$$

$$\textit{où } \beta = g(\alpha e^{At_{n-k}}, C_{n-k}, \ldots, (C_n, t_n))$$

Démonstration. On démontre ce résultat par récurrence sur le nombre de sous-ensembles C_j. S'il n'y a qu'un sous-ensemble en présence, la propriété est démontrée dans le Lemme 4.3. Supposons que la relation soit vraie pour l sous-ensembles C_j, $l \leq k$.

$$\begin{aligned}
& P_\alpha(X_{t_{n+1}} \in B / X_{t_n} \in C_n, \ldots, X_{t_{n-k}} \in C_{n-k}) \\
= \;& \frac{P_\alpha(X_{t_{n+1}} \in B, X_{t_n} \in C_n, \ldots, X_{t_{n-k}} \in C_{n-k})}{P_\alpha(X_{t_n} \in C_n, \ldots, X_{t_{n-k}} \in C_{n-k})} \\
= \;& \frac{\sum_{c \in C_n} P_\alpha(X_{t_{n+1}} \in B / X_{t_n} = c) P_\alpha(X_{t_n} = c / X_{t_{n-1}} \in C_{n-1}, \ldots, X_{t_{n-k}} \in C_{n-k})}{\sum_{c \in C_n} P_\alpha(X_{t_n} = c / X_{t_{n-1}} \in C_{n-1}, \ldots, X_{t_{n-k}} \in C_{n-k})} \\
= \;& \sum_{c \in C_n} e^{A(t_{n+1}-t_n)}(c, B) \left(\frac{P_\gamma(X_{t_n - t_{n-1}} = c)}{P_\gamma(X_{t_n - t_{n-1}} \in C_n)} \right) \quad \text{par hypothèse de récurrence} \\
& \text{où } \gamma = g(\alpha e^{At_{n-k}}, C_{n-k}, \ldots, (C_{n-1}, t_{n-1})) \\
= \;& \sum_{c \in C_n} e^{A(t_{n+1}-t_n)}(c, B) \left(\frac{P_{\gamma e^{A(t_n - t_{n-1})}}(X_0 = c)}{P_{\gamma e^{A(t_n - t_{n-1})}}(X_0 \in C_n)} \right) \quad \text{par le Lemme 4.1} \\
= \;& \sum_{c \in C_n} (\gamma e^{A(t_n - t_{n-1})})^{C_n}(c) e^{A(t_{n+1}-t_n)}(c, B) \\
= \;& P_\beta(X_{t_{n+1}-t_n} \in B) \quad \text{puisque } (\gamma e^{A(t_n - t_{n-1})})^{C_n} = \beta
\end{aligned}$$

□

Lemme 4.9 *Le vecteur $\beta = g(\alpha, C_0, (C_1, t_1), \ldots, (C_j, t_j))$ est la distribution de X_{t_j} sachant $\{X_0 \in C_0, \ldots, X_{t_j} \in C_j\}$. C'est à dire:*

$$\beta(i) = P_\alpha(X_{t_j} = i / X_0 \in C_0, \ldots, X_{t_j} \in C_j)$$

Démonstration. Si $i \notin C_j$, la relation est triviale. Supposons donc que $i \in C_j$.

$$\begin{aligned}
& P_\alpha(X_{t_j} = i / X_0 \in C_0, \ldots, X_{t_j} \in C_j) \\
= \;& \frac{P_\alpha(X_{t_j} = i, X_0 \in C_0, \ldots, X_{t_j} \in C_j)}{P_\alpha(X_0 \in C_0, \ldots, X_{t_j} \in C_j)} \\
= \;& \frac{P_\alpha(X_{t_j} = i, X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1})}{P_\alpha(X_0 \in C_0, \ldots, X_{t_j} \in C_j)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{P_\alpha(X_{t_j} = i / X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1}) P_\alpha(X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1})}{P_\alpha(X_{t_j} \in C_j / X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1}) P_\alpha(X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1})} \\
&= \frac{P_\alpha(X_{t_j} = i / X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1})}{P_\alpha(X_{t_j} \in C_j / X_0 \in C_0, \ldots, X_{t_{j-1}} \in C_{j-1})} \\
&= \frac{P_\delta(X_{t_j - t_{j-1}} = i)}{P_\delta(X_{t_j - t_{j-1}} \in C_j)} \quad \text{où } \delta = g(\alpha, C_0, \ldots, (C_{j-1}, t_{j-1})), \text{ par le Lemme 4.8} \\
&= (\delta e^{A(t_j - t_{j-1})})^{C_j}(i) = \beta(i)
\end{aligned}
$$

□

Lemme 4.10 *La fonction* $v \mapsto P_v(X_{t_j} \in B / X_{t_{j-1}} \in C_{j-1}, \ldots, X_{t_{j-k}} \in C_{j-k})$ *pour* j, k, $C_{j-1}, \ldots, C_{j-k}$, B, $0 < t_{j-k} < \cdots < t_j$ *fixés, est continue sur* $\mathcal{A}$.

Démonstration. La démonstration est identique à celle du Lemme 3.8. □

4.2 Relation avec le cas discret. Uniformisation

Comme nous l'avons fait dans le cas discret, nous allons étudier maintenant le processus $Y = agg(\alpha, A, \mathcal{B})$. On conserve les mêmes définitions pour les notions *"fortement agrégeable"* (ou *"agrégeable"*) et *"faiblement agrégeable"*.

La notion d'agrégation forte étant un cas particulier de celle d'agrégation faible, on ne parlera dans cette section que du cas général.

On notera par $\mathcal{C}(\alpha, \mathcal{B})$ l'ensemble suivant:

$\mathcal{C}(\alpha, B) \stackrel{\text{déf}}{=} \{\ \beta \in \mathcal{A}\ /\ \exists j \geq 0$ et une suite $(C_1, \ldots, C_j, B)$ possible pour α, réduite à (B) si $j = 0$, tels que $\beta = g(\alpha, C_1, (C_2, t_1), \ldots, (C_j, t_{j-1}), (B, t_j))$, pour tout $0 = t_0 < t_1 < \cdots < t_j\ \}$

Cet ensemble est l'équivalent de l'ensemble $\mathcal{A}(\alpha, B)$ dans le cas discret.

On peut maintenant énoncer le théorème suivant qui est la version pour les processus à temps continu du Théorème 3.9.

Théorème 4.11 *Le processus* $Y = agg(\alpha, A, \mathcal{B})$ *est un processus markovien homogène si et seulement si* $\forall l, m \in F$ *et* $\forall t \geq 0$*, la probabilité* $P_\beta(X_t \in B(m))$ *a la même valeur pour tout* $\beta \in \mathcal{C}(\alpha, B(l))$*. Cette valeur commune est la probabilité que le processus* Y *soit dans l'état* m *à l'instant* t *sachant qu'il est dans l'état* l *à l'instant* 0.

Démonstration. La démonstration est basée sur le Lemme 4.8. Ce lemme nous permet d'écrire:

$\forall l, m \in F$, $\forall n \geq 0$, $\forall (C_0, C_1, \ldots, C_{n-1}, B(l))$ possible pour α et $\forall\ 0 < t_1 < \cdots < t_n$,

$$P_\alpha(X_{t_n+t} \in B(m) / X_{t_n} \in B(l),\ X_{t_{n-1}} \in C_{n-1}, \ldots, X_0 \in C_0) = P_\beta(X_t \in B(m))$$

$$\text{où } \beta = g\,(\alpha,\ C_0,\ (C_1, t_1),\ \ldots, (C_{n-1}, t_{n-1}),\ (B(l), t_n))$$

Si la condition du théorème est satisfaite, la quantité ci-dessus ne dépend que de l, m et t, et ceci est équivalent au fait que Y est un processus markovien homogène.

Réciproquement, supposons que Y est un processus markovien homogène. Soit β un vecteur quelconque de $\mathcal{C}(\alpha, B(l))$. Le vecteur β peut donc s'écrire:

$$\beta = g(\alpha,\ C_0,\ (C_1, t_1),\ \ldots, (C_{n-1}, t_{n-1}),\ (B(l), t_n))$$

D'où, en réutilisant le Lemme 4.8,

$$\begin{aligned} P_\beta(X_t \in B(m)) &= P_\alpha(X_{t_n+t} \in B(m) / X_{t_n} \in B(l),\ X_{t_{n-1}} \in C_{n-1}, \ldots, X_{t_0} \in C_0) \\ &= P_\alpha(X_{t_n+t} \in B(m) / X_{t_n} \in B(l)) \end{aligned}$$

qui ne dépend pas de t_n par homogénéité de Y, donc, $P_\beta(X_t \in B(m))$ a la même valeur pour tout $\beta \in \mathcal{C}(\alpha, B(l))$. □

Ce théorème est important pour la preuve du suivant qui est à la base de toute cette section.

Théorème 4.12 *Le processus $Y = agg(\alpha, A, \mathcal{B})$ est un processus markovien homogène si et seulement si $\forall l, m \in F$, la quantité $\sum_{i \in B(l)} \beta(i) A(i, B(m))$ est la même pour tout $\beta \in \mathcal{C}(\alpha, B(l))$. Cette valeur commune est le taux de transition de l'état l vers l'état m pour Y lorsque $l \neq m$.*

Démonstration. Pour prouver ce théorème, il est suffisant de prouver que la condition citée est équivalente à celle citée dans le Théorème 4.11. On donne d'abord deux résultats simples:

$$\forall l, m \in F$$

$$\lim_{t \to 0} \frac{1}{t} \sum_{i \in B(l)} \beta(i) e^{At}(i, B(m)) = \sum_{i \in B(l)} \beta(i) A(i, B(m)) \text{ si } l \neq m$$

$$\lim_{t \to 0} \frac{1}{t} \left(\sum_{i \in B(l)} \beta(i) e^{At}(i, B(l)) - 1 \right) = \sum_{i \in B(l)} \beta(i) A(i, B(l))$$

Supposons que la condition du Théorème 4.11 soit vraie. Alors, grâce à ces deux résultats, la condition du Théorème 4.12 est aussi vraie.

Pour prouver la réciproque, il suffit de prouver que la condition du Théorème 4.12 implique la condition suivante:

$\forall l, m \in F$ et $\forall k \geq 0$, $\sum_{i \in B(l)} \beta(i) A^k(i, B(m))$ a la même valeur pour tout $\beta \in \mathcal{C}(\alpha, B(l))$, puisque:

$$P_\beta(X_t \in B(m)) = \sum_{i \in B(l)} \beta(i) e^{At}(i, B(m)) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{i \in B(l)} \beta(i) A^k(i, B(m))$$

Pour $k = 0$ et pour $k = 1$ c'est trivial.

Supposons que l'implication soit vraie jusqu'à l'ordre $k-1$. On a donc: $\forall l, m \in F$, $\sum_{i \in B(l)} \beta(i) A^{k-1}(i, B(m))$ a la même valeur pour tout $\beta \in \mathcal{C}(\alpha, B(l))$. Soit $\beta \in \mathcal{C}(\alpha, B(l))$: $\beta = g(\alpha,\ C_0,\ (C_1, t_1),\ \ldots, (C_{n-1}, t_{n-1}),\ (B(l), t_n))$.

$$\sum_{i \in B(l)} \beta(i) A^k(i, B(m)) = \Sigma_1 + \Sigma_2 \ \text{ avec:}$$

$$\Sigma_1 = \sum_{h/h \neq l} \sum_{j \in B(h)} A^{k-1}(j, B(m)) \sum_{i \in B(l)} \beta(i) A(i, j)$$

$$\text{et } \ \Sigma_2 = \sum_{j \in B(l)} A^{k-1}(j, B(m)) \sum_{i \in B(l)} \beta(i) A(i, j)$$

Soit $\tau > t_n$ et $h \in F$ tel que $(B(h))$ est possible pour β. On définit:

$$\gamma_h^\tau \stackrel{\text{déf}}{=} g\left(\beta, (B(h), \tau)\right) = \left(\beta e^{A(\tau - t_n)}\right)^{B(h)}$$

On a alors:

$$\gamma_h^\tau(j) = \begin{cases} \dfrac{\sum_{i \in B(l)} \beta(i) e^{A(\tau - t_n)}(i, j)}{\sum_{i \in B(l)} \beta(i) e^{A(\tau - t_n)}(i, B(h))} & \text{si } j \in B(h) \\ 0 & \text{sinon} \end{cases}$$

$$\begin{aligned} \lim_{\tau \to t_n} \gamma_h^\tau(j) &= \frac{\lim_{\tau \to t_n} \frac{\sum_{i \in B(l)} \beta(i) e^{A(\tau - t_n)}(i, j)}{\tau - t_n}}{\frac{\sum_{i \in B(l)} \beta(i) e^{A(\tau - t_n)}(i, B(h))}{\tau - t_n}} \\ &= \frac{\sum_{i \in B(l)} \beta(i) A(i, j)}{\sum_{i \in B(l)} \beta(i) A(i, B(h))} \quad \text{si } h \neq l \end{aligned}$$

Donc, pour $h \neq l$, on a:

$$\sum_{i \in B(l)} \beta(i) A(i, j) = \left(\sum_{i \in B(l)} \beta(i) A(i, B(h)) \right) \lim_{\tau \to t_n} \gamma_h^\tau(j)$$

En remplaçant maintenant cette valeur dans Σ_1, on obtient:

$$\Sigma_1 = \sum_{h/h\neq l} \left(\sum_{i\in B(l)} \beta(i)A(i,B(h)) \right) \lim_{\tau\to t_n} \sum_{j\in B(h)} \gamma_h^\tau(j)A^{k-1}(j,B(m))$$

$\sum_{i\in B(l)} \beta(i)A(i,B(h))$ a la même valeur pour tout $\beta \in C(\alpha, B(l))$ par hypothèse. La quantité $\sum_{j\in B(h)} \gamma_h^\tau(j)A^{k-1}(j,B(m))$ a la même valeur pour tout $\gamma_h^\tau \in C(\alpha, B(h))$ par hypothèse de récurrence et donc elle a aussi la même valeur pour tout $\beta \in C(\alpha, B(l))$ par définition de γ_h^τ. On en conclut que Σ_1 a la même valeur pour tout $\beta \in C(\alpha, B(l))$.

Soit maintenant $j \in B(l)$, on a par définition:

$$\gamma_l^\tau(j) = \frac{\sum_{i\in B(l)} \beta(i)e^{A(\tau-t_n)}(i,j)}{\sum_{i\in B(l)} \beta(i)e^{A(\tau-t_n)}(i,B(l))}$$

qui peut s'écrire de la façon suivante:

$$\gamma_l^\tau(j) = \frac{\beta(j) + \sum_{i\in B(l)} \beta(i)(e^{A(\tau-t_n)} - I)(i,j)}{1 + \sum_{i\in B(l)} \beta(i)(e^{A(\tau-t_n)} - I)(i,B(l))}$$

Il s'en suit immédiatement:

$$\lim_{\tau\to t_n} \frac{\gamma_l^\tau(j) - \beta(j)}{\tau - t_n} = \sum_{i\in B(l)} \beta(i)A(i,j) - \sum_{i\in B(l)} \beta(i)A(i,B(l))$$

En tirant de cette expression la quantité $\sum_{i\in B(l)} \beta(i)A(i,j)$ puis en la reportant dans Σ_2, on obtient:

$$\Sigma_2 = \sum_{j\in B(l)} A^{k-1}(j,B(m)) \left[\beta(j) \sum_{i\in B(l)} \beta(i)A(i,B(l)) + \lim_{\tau\to t_n} \frac{\gamma_l^\tau(j) - \beta(j)}{\tau - t_n} \right]$$

qui peut encore s'écrire

$$\begin{aligned} \Sigma_2 &= \left(\sum_{i\in B(l)} \beta(i)A(i,B(l)) \right) \left(\sum_{j\in B(l)} \beta(j)A^{k-1}(j,B(m)) \right) \\ &+ \lim_{\tau\to t_n} \frac{1}{\tau - t_n} \left[\sum_{j\in B(l)} \gamma_l^\tau(j)A^{k-1}(j,B(m)) - \sum_{j\in B(l)} \beta(j)A^{k-1}(j,B(m)) \right] \end{aligned}$$

Chacune des quatre sommes en présence a la même valeur pour tout $\beta \in C(\alpha, B(l))$ par hypothèse, hypothèse de récurrence et par définition de γ_l^τ. La quantité Σ_2 a donc aussi la même valeur pour tout $\beta \in C(\alpha, B(l))$, ce qui termine la démonstration. □

On notera $\mathcal{C}_{\mathcal{M}}$ l'ensemble des distributions initiales α conduisant à un processus markovien homogène, c'est à dire:

$$\mathcal{C}_{\mathcal{M}} \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A} / Y = agg(\alpha, A, \mathcal{B}) \quad \text{est un processus markovien homogène}\}$$

En conservant la même démarche que pour le cas discret, on obtient les résultats suivants.

Théorème 4.13 *Si $\alpha \in \mathcal{C}_{\mathcal{M}}$ alors $\alpha e^{As} \in \mathcal{C}_{\mathcal{M}}$ pour tout $s \geq 0$*

Démonstration. Soit $\alpha \in \mathcal{C}_{\mathcal{M}}$. En utilisant le Lemme 4.8, il est simple de vérifier que:

$$P_{\alpha}(X_{t_{n+1}+s} \in B / X_{t_n+s} \in D,\ X_{t_{n-1}+s} \in C_{n-1},\ \ldots,\ X_s \in C_0)$$

$$= P_{\alpha e^{As}}(X_{t_{n+1}} \in B / X_{t_n} \in D,\ X_{t_{n-1}} \in C_{n-1},\ \ldots,\ X_0 \in C_0)$$

Par hypothèse, la partie gauche cette égalité est indépendante de n, de s et de la suite $(C_0,\ C_1,\ \ldots,\ C_{n-1})$ considérée (les classes B et D étant fixées). Donc, $\alpha e^{As} \in \mathcal{C}_{\mathcal{M}}$ pour tout $s \geq 0$. □

Corollaire 4.14 *Si $\mathcal{C}_{\mathcal{M}} \neq \emptyset$ et si $\hat{A}$ désigne la matrice des taux de transition du processus markovien homogène agrégé $Y = agg(\alpha, A, \mathcal{B})$ alors $\hat{A}$ est la même pour tout α conduisant à un processus markovien homogène agrégé et de plus $\pi \in \mathcal{C}_{\mathcal{M}}$*

Démonstration. Soit $\alpha \in \mathcal{A}$ tel que $Y = agg(\alpha, A, \mathcal{B})$ soit un processus markovien homogène de matrice des taux de transition $\hat{A}$. On a alors:

$$\begin{aligned} e^{\hat{A}s}(l,m) &= P_{\alpha}(X_{t+s} \in B(m) / X_t \in B(l)) \text{ pour tout } t \geq 0 \\ &= P_{\alpha e^{At}}(X_s \in B(m) / X_0 \in B(l)) \text{ pour tout } t \geq 0 \end{aligned}$$

En faisant tendre t vers $+\infty$, on obtient grâce au Lemme 4.10:

$$e^{\hat{A}s}(l,m) = P_{\pi}(X_s \in B(m) / X_0 \in B(l))$$

Ce qui donne:

$$\hat{A}(l,m) = \lim_{s \to 0} \frac{1}{s} P_{\pi}(X_s \in B(m) / X_0 \in B(l)) \text{ pour } l \neq m$$

$$\hat{A}(l,l) = \lim_{s \to 0} \frac{1}{s} \left(P_{\pi}(X_s \in B(m) / X_0 \in B(l)) - 1\right)$$

$\square$

On rappelle que $(.\,, A)$ est *faiblement agrégeable* par rapport à la partition $\mathcal{B}$ si et seulement si $\mathcal{C}_{\mathcal{M}} \neq \emptyset$. Dans ce cas, on notera $\hat{A}$ la matrice des taux de transition de $agg(\alpha, A, \mathcal{B})$ pour tout $\alpha \in \mathcal{C}_{\mathcal{M}}$.

De même que dans le cas discret, on introduit les notations suivantes.

- pour $i \in F$, on notera $\tilde{A}_i$ la matrice $(n(i), M)$ définie par: $\tilde{A}_i(j,k) \stackrel{\text{déf}}{=} A(n(1)+\cdots+n(i-1)+j, B(k))$, $1 \leq j \leq n(i)$, $k \in F$

On en déduit: $\hat{A}_j = (T_j.\pi^{B(j)})\tilde{A}_j$ où $\hat{A}_j$ désigne la $j^{ième}$ ligne de $\hat{A}$. Dans la suite, comme dans le cas discret, on définira toujours $\hat{A}$ par la relation précédente même lorsque Y n'est pas markovien homogène.

- pour $j \in F$, on notera δ_j le système linéaire $x_j\tilde{A}_j = \hat{A}_j$ à M équations et $n(j)$ inconnues.
- $\mathcal{C}^* \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A} \,/\, T_l.\alpha^{B(l)}$ est solution de δ_l pour tout $l \in F$ tel que $\alpha_{B(l)} \neq 0\}$.

Comme dans le cas discret, on donne une caractérisation de l'ensemble $\mathcal{C}_{\mathcal{M}}$ en définissant pour tout $j \geq 1$:

$$\mathcal{C}^j \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{C}^* \;/\; \forall \beta = g(\alpha,\, B(i_1),\, \ldots, (B(i_k), t_{i_k})),\; k \leq j,\; \beta \in \mathcal{C}^*\}$$

Avec cette notation, on a:

$$\mathcal{C}^1 = \{\alpha \in \mathcal{C}^* \;/\; \forall m \in F,\; \alpha^{B(m)} \in \mathcal{C}^*\} = \mathcal{C}^*$$

et on obtient facilement:

$$\forall j \geq 1,\; \mathcal{C}^{j+1} \subseteq \mathcal{C}^j$$

ainsi que l'équivalent suivant du Théorème 3.20.

Théorème 4.15

$$\mathcal{C}_{\mathcal{M}} = \bigcap_{j \geq 1} \mathcal{C}^j$$

Démonstration. En utilisant le Théorème 4.12, la démonstration est quasiment identique à celle du Théorème 3.20. □

Les mêmes démonstrations que dans le cas discret servent pour prouver que:

$$\forall j \geq 1,\ \mathcal{C}^{j+1} = \{\alpha \in \mathcal{C}^j \ / \ \forall t \geq 0\ ,\ \alpha^{B(l)} e^{At} \in \mathcal{C}^j \forall l \in F \text{ tel que } \alpha_{B(l)} \neq 0\}$$

Pour terminer l'étude du cas continu, on va montrer que l'on peut se ramener au cas discret par uniformisation du processus X.

La technique d'uniformisation [5] consiste à construire une chaîne de Markov $Z = \{Z_n, n \in \mathbb{N}\}$ sur l'espace d'états E, de matrice des probabilités de transition P, et un processus de Poisson $\{N(t), t \geq 0\}$ indépendant de Z et de taux λ tels que $X_t = Z_{N(t)}$, de la façon suivante:

- on choisit $\lambda \in \mathbb{R}$ tel que: $\lambda \geq max(\lambda(i), i = 1, ..., N)$
- on pose $P(Z_{n+1} = j / Z_n = i) = P(i,j) = \left\{ \begin{array}{ll} 1 - \frac{\lambda(i)}{\lambda} & si\ i = j \\ \frac{A(i,j)}{\lambda} & sinon \end{array} \right.$

On obtient alors la relation suivante liant les matrices A et P.

$$P = I + A/\lambda \tag{2}$$

où I est la matrice unité de dimension N.

On notera W la chaîne agrégée de Z par rapport à la partition $\mathcal{B}$. On introduit maintenant les notations suivantes. Soit une distribution de probabilité initiale $\alpha \in \mathcal{A}$, une suite $(C_0, C_1, \ldots, C_j)$ de sous-ensembles de E et des entiers naturels $0 = n_0 < n_1 < \ldots < n_j$ tels que $P_\alpha(X_{n_0} \in C_0, X_{n_1} \in C_1, \ldots, X_{n_j} \in C_j) > 0$, on définit alors le vecteur $h(\alpha, C_0, (C_1, n_1), \ldots, (C_j, n_j))$ récursivement par:

$$\begin{aligned} h(\alpha, C) &= \alpha^C \\ h(\alpha, C_0, (C_1, n_1)) &= (h(\alpha, C) P^{n_1})^{C_1} \\ \text{et pour } k \geq 2 : & \\ h(\alpha, C_0, (C_1, n_1), \ldots, (C_k, n_k)) &= \left(h(\alpha, C_0, (C_1, n_1), \ldots, (C_{k-1}, n_{k-1})) P^{n_k - n_{k-1}}\right)^{C_k} \end{aligned}$$

Il est immédiat de vérifier que cette fonction satisfait des propriétés analogues à celles des fonctions f et g. Signalons simplement les deux suivantes:

- (i) La restriction de $\beta = h(\alpha, C_0, (C_1, n_1), \ldots, (C_k, n_k))$ à C_k est:

$$\beta_{C_k} = K^{-1} \alpha_{C_0} P_{C_0 C_1}^{n_1} P_{C_1 C_2}^{n_2 - n_1} \ldots P_{C_{n_{k-1}} C_{n_k}}^{n_k - n_{k-1}}$$

où K est la constante de normalisation:

$$K = \alpha_{C_0} P^{n_1}_{C_0 C_1} P^{n_2 - n_1}_{C_1 C_2} \ldots P^{n_k - n_{k-1}}_{C_{n_{k-1}} C_{n_k}} 1^T$$

- (ii) Si la suite $(C_0, C_1, \ldots, C_j)$ est possible pour $\alpha_1, \ldots, \alpha_R$ alors pour tous $\lambda_1, \ldots, \lambda_R$ tels que $\lambda_r \geq 0$ et $\sum_{1 \leq r \leq R} \lambda_r = 1$, elle est aussi possible pour le vecteur $\sum_{1 \leq r \leq R} \lambda_r \alpha_r$ et l'on a:

$$h\left(\sum_{1 \leq r \leq R} \lambda_r \alpha_r, C_0, \ldots, (C_k, n_k)\right) = \sum_{1 \leq r \leq R} \lambda_r \frac{K_r}{K} h(\alpha_r, C_0, \ldots, (C_k, n_k))$$

$$\text{où } K_r = (\alpha_r)_{C_0} P^{n_1}_{C_0 C_1} P^{n_2 - n_1}_{C_1 C_2} \ldots P^{n_k - n_{k-1}}_{C_{n_{k-1}} C_{n_k}} 1^T$$

$$\text{et } K = \sum_{1 \leq r \leq R} \lambda_r K_r$$

Les preuves sont identiques à celles données pour les fonctions f et g.

On notera aussi $\mathcal{D}(\alpha, \mathcal{B})$ l'ensemble suivant:

$\mathcal{D}(\alpha, B) \stackrel{\text{déf}}{=} \{ \beta \in \mathcal{A} \ / \ \exists j \geq 0$ et une suite $(C_1, \ldots, C_j, B)$, réduite à (B) si $j = 0$, tels que $\beta = h(\alpha, C_1, (C_2, t_1), \ldots, (C_j, n_{j-1}), (B, n_j))$, pour tout $0 = n_0 < n_1 < \cdots < n_j \}$

Une autre version du Théorème 3.9 est alors:

Théorème 4.16 *La chaîne $W = agg(\alpha, P, \mathcal{B})$ est une chaîne de Markov homogène si et seulement si $\forall l, m \in F$, la probabilité $P_\beta(X_1 \in B(m))$ a la même valeur pour tout $\beta \in \mathcal{D}(\alpha, B(l))$. Cette valeur commune est la probabilité de transition de l'état l vers l'état m pour la chaîne W.*

Démonstration. La démonstration est identique à celles des Théorèmes 3.9 et 4.11 □

Remarque On considère ici la fonction h plutôt que la fonction f utilisée dans le cas discret, pour que le raisonnement soit plus simple. Rappelons que dans cette section, on cherche un lien entre les cas discret et continu. A partir de cette fonction h, on va caractériser l'ensemble $\mathcal{A}_\mathcal{M}$ des distributions initiales de la chaîne de Markov homogène Z conduisant à une chaîne de Markov homogène pour l'agrégé W, d'une manière analogue à celle utilisée dans le cas discret et qui est la suivante.

Rappelons tout d'abord que:

$$\mathcal{A}_\mathcal{M} = \{\alpha \in \mathcal{A} / W = agg(\alpha, P, \mathcal{B}) \text{ est une chaîne de Markov homogène}\}$$

On note pour $j \geq 1$:

- $\mathcal{D}^j \stackrel{\text{déf}}{=} \{\alpha \in \mathcal{A}^* \;/\; \forall \beta = h(\alpha, B(i_1), \ldots, (B(i_k), n_{i_k})),\; k \leq j,\; \beta \in \mathcal{A}^*\}$. On peut alors remarquer que:
  - $\mathcal{D}^1 = \mathcal{A}^* = \mathcal{A}^1$
  - par contre, pour $j > 1$, on n'a pas en général $\mathcal{D}^j = \mathcal{A}^j$

De même que dans les cas discret et continu, on a les résultats suivants que l'on donne sans démonstration car elles sont analogues à celles du Lemme 3.21 et du Théorème 3.34:

- $\forall j \geq 1$,

  $\mathcal{D}^{j+1} = \{\alpha \in \mathcal{D}^j \;/\; \forall n \in \mathbb{N},\; \alpha^{B(l)} P^n \in \mathcal{D}^j$ pour tout $l \in F$ tel que $\alpha_{B(l)} \neq 0\}$

- $\mathcal{A}_{\mathcal{M}} = \bigcap_{j \geq 1} \mathcal{D}^j = \bigcap_{j \geq 1} \mathcal{A}^j$

Pour plus de clarté, on introduit les notations suivantes:

- pour tout $i \in F$, la matrice $\tilde{I}_l$ de taille $(n(l), M)$ définie par:

$$\tilde{I}_l(i,j) = \begin{cases} 1 & \text{si } j = l \\ 0 & \text{sinon} \end{cases}$$

- e_l le vecteur ligne de taille M dont la $l^{ième}$ composante vaut 1 et toutes les autres 0.

- la matrice identité (M, M) que l'on note $\hat{I}$.

On peut maintenant énoncer les lemmes suivants:

Lemme 4.17

$$\hat{P} = \hat{I} + \hat{A}/\lambda$$

Démonstration. Soit $l \in F$,

$$\begin{aligned} \hat{A}_l &= T_l.\pi^{B(l)} \tilde{A}_l \\ &= T_l.\pi^{B(l)} \lambda(\tilde{P}_l - \tilde{I}_l) \text{ par la relation (2).} \end{aligned}$$

Ce qui nous donne en développant:

$$\hat{A}_l = \lambda(\hat{P}_l - e_l)$$

c'est à dire:

$$\hat{A} = -\lambda(I - \hat{P})$$

d'où le résultat. □

Lemme 4.18 $\mathcal{D}^1 = \mathcal{C}^1$

Démonstration. Soit $\alpha \in \mathcal{A}$,

$$\begin{aligned}
\alpha \in \mathcal{D}^1 &\iff \alpha \in \mathcal{A}^* \\
&\iff T_l.\alpha^{B(l)}\tilde{P}_l = \hat{P}_l \;\; \forall l \in F \\
&\iff T_l.\alpha^{B(l)}(\tilde{I}_l + \tilde{A}_l/\lambda) = e_l + \hat{A}_l/\lambda \;\; \forall l \in F \text{ par le lemme précédent} \\
&\iff T_l.\alpha^{B(l)}\tilde{A}_l = \hat{A}_l \;\; \forall l \in F \\
&\iff \alpha \in \mathcal{C}^* \\
&\iff \alpha \in \mathcal{C}^1
\end{aligned}$$

d'où, $\mathcal{D}^1 = \mathcal{C}^1$. □

On peut maintenant énoncer le théorème principal du cas continu.

Théorème 4.19

$$\mathcal{A}_{\mathcal{M}} = \mathcal{C}_{\mathcal{M}}$$

Démonstration. Il nous suffit de prouver que pour tout $j \geq 1$, $\mathcal{D}^j = \mathcal{C}^j$.

On vient de prouver que $\mathcal{D}^1 = \mathcal{C}^1$. Supposons maintenant que $\mathcal{D}^j = \mathcal{C}^j$ pour un entier $j > 1$ fixé. Rappelons que:

$$\mathcal{C}^{j+1} = \{\alpha \in \mathcal{C}^j / \; \forall t \geq 0, \; \alpha^{B(l)}e^{At} \in \mathcal{C}^j, \text{ pour tout } l \in F \text{ tel que } \alpha_{B(l)} \neq 0\}$$

$$\mathcal{D}^{j+1} = \{\alpha \in \mathcal{D}^j \; / \; \forall n \in \mathbb{N} \, , \; \alpha^{B(l)}P^n \in \mathcal{D}^j \text{ pour tout } l \in F \text{ tel que } \alpha_{B(l)} \neq 0\}$$

L'hypothèse de récurrence nous permet d'écrire:

$$\mathcal{C}^{j+1} = \{\alpha \in \mathcal{D}^j / \; \forall t \geq 0, \; \alpha^{B(l)}e^{At} \in \mathcal{D}^j, \text{ pour tout } \alpha \text{ tel que } \alpha_{B(l)} \neq 0\}$$

Remarquons tout d'abord que l'on a la propriété suivante:

$$\forall t \geq 0, \quad e^{At} = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}P^n$$

D'où:

$$\forall t \geq 0, \;\; \forall l \in F \text{ tel que } \alpha_{B(l)} \neq 0, \quad \alpha^{B(l)}e^{At} = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\alpha^{B(l)}P^n$$

Soit $\alpha \in \mathcal{C}^{j+1}$, c'est à dire soit $\alpha \in \mathcal{D}^j$ tel que $\forall t \geq 0$, $\alpha^{B(l)}e^{At} \in \mathcal{D}^j$ pour tout α tel que $\alpha_{B(l)} \neq 0$.

$$\alpha^{B(l)}e^{At} \in \mathcal{D}^j \iff h(\alpha^{B(l)}e^{At}, B(i_1), \ldots, (B(i_k), n_k)) \in \mathcal{A}^* \ \forall k \leq j$$

Or,

$$\begin{aligned} & h\left(\alpha^{B(l)}e^{At}, B(i_1), \ldots, (B(i_k), n_k)\right) \\ &= h\left(\sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\alpha^{B(l)}P^n, B(i_1), \ldots, (B(i_k), n_k)\right) \\ &= \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\frac{K_n}{K}h\left(\alpha^{B(l)}P^n, B(i_1), \ldots, (B(i_k), n_k)\right) \end{aligned}$$

$$\text{où } K_n = (\alpha^{B(l)}P^n)_{B(i_1)}P^{n_2}_{B(i_1),B(i_2)}P^{n_3-n_2}_{B(i_2),B(i_3)}\cdots P^{n_k-n_{k-1}}_{B(i_{k-1})B(i_k)}1^T$$

$$\text{et } K = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}K_n$$

Cette dernière égalité résulte des propriétés (i) et (ii) de la fonction h (page 41), en vérifiant que (ii) reste valable pour une somme infinie.

Finalement, on obtient:

$$\alpha^{B(l)}e^{At} \in \mathcal{D}^j \iff \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\frac{K_n}{K}T_{i_k}.h(\alpha^{B(l)}P^n, B(i_1), \ldots, (B(i_k), n_k))\tilde{P}_{i_k} = \hat{P}_{i_k}$$

$$\iff \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\frac{K_n}{K}\left(T_{i_k}.h(\alpha^{B(l)}P^n, B(i_1), \ldots, (B(i_k), n_k))\tilde{P}_{i_k} - \hat{P}_{i_k}\right) = 0$$

Cette dernière égalité devant avoir lieu pout tout $t \geq 0$, on en déduit simplement:

$$\begin{aligned} \alpha^{B(l)}e^{At} \in \mathcal{D}^j &\iff T_{i_k}.h(\alpha^{B(l)}P^n, B(i_1), \ldots, (B(i_k), n_k))\tilde{P}_{i_k} = \hat{P}_{i_k} \\ &\iff h(\alpha^{B(l)}P^n, B(i_1), \ldots, (B(i_k), n_k)) \in \mathcal{A}^* \\ &\iff \alpha^{B(l)}P^n \in \mathcal{D}^j \\ &\iff \alpha \in \mathcal{D}^{j+1} \end{aligned}$$

Par cette suite d'équivalences, on a prouvé que $\mathcal{C}^{j+1} = \mathcal{D}^{j+1}$. La récurrence permet donc de conclure que $\forall j \geq 1, \mathcal{C}^j = \mathcal{D}^j$ èt donc $\mathcal{A}_{\mathcal{M}} = \mathcal{C}_{\mathcal{M}}$. □

Remarque Dans [2], on peut trouver une preuve de l'équivalence entre l'agrégation forte et la condition sur A correspondante à celle du Théorème 3.15, bien que le fait que la propriété markovienne soit vérifiée par tous les processus partageant le même générateur infinitesimal ne soit pas clairement signalée. Ce résultat n'est en fait qu'un corollaire immédiat du théorème précédent et du Théorème 3.15:

Corollaire 4.20 *$(\,.\,,A)$ est fortement agrégeable par rapport à la partition $\mathcal{B}$ si et seulement si pour toute paire de sous-ensembles $D, B \in \mathcal{B}$, $A(d,B)$ a la même valeur pour tout $d \in D$. Cette valeur commune est le taux de transition de D vers B pour le processus Y si $D \neq B$.*

5 Conclusions

Le but de ce travail est la caractérisation des chaînes ou des processus de Markov faiblement agrégeables par rapport à une partition fixée de l'espace d'état.

Dans le cas discret, le résultat essentiel est donné dans l'algorithme de la fin de la Section 3 ou dans le Théorème 3.34. La caractérisation, donnée dans [6], de l'ensemble des distributions initiales conduisant à un processus agrégé markovien homogène sous la forme d'une intersection infinie d'ensembles est ramenée à une intersection finie d'au plus N ensembles où N est la taille de la chaîne d'origine.

Dans le cas continu, on a montré qu'un processus markovien homogène X est faiblement agrégeable par rapport à une partition fixée de l'espace d'état si et seulement si la chaîne de Markov uniformisée associée l'est. De plus, l'ensemble des distributions initiales de X conduisant à un processus agrégé markovien homogène est identique à l'ensemble des distributions initiales de la chaîne de Markov uniformisée conduisant à une chaîne agrégée de Markov homogène. Ceci signifie que dans la pratique, pour déterminer cet ensemble, on pourra construire d'abord la chaîne de Markov uniformisée du processus, puis déterminer l'ensemble correspondant à cette chaîne à partir de l'étude faite dans le cas discret.

En ce qui concerne les perspectives de continuation de cette étude, il semble évident que la recherche d'applications pour lesquelles les processus de Markov concernés soient faiblement agrégeables est l'une des premières tâches possibles. D'autre part, le cas où l'espace d'état est infini mérite une attention particulière, en distinguant éventuellement des partitions finies ou infinies composées de classes elles-mêmes finies ou infinies.

Bibliographie

[1] A.M. Abdel-Moneim and F.W. Leysieffer. Weak lumpability in finite Markov chains. *J.Appl.Prob.*, 19:685–691, 1982.

[2] J.M.Ferrandiz and A.A.Lazar. *Geometric Aggregation, Expansion and Partial Observation for Markov Processes.* Technical Report, Columbia University, New York, NY 10027, 1986.

[3] J.G. Kemeny and J.L Snell. *Finite Markov chains.* Springer-Verlag, New York Heidelberg Berlin, 1976.

[4] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, New-Jersey, 1970.

[5] S.M. Ross. *Stochastic Processes.* John Wiley and Sons, 1983.

[6] G. Rubino and B. Sericola. *On weak lumpability in Markov chains.* Technical Report 387, IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, 1987.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique